Author:
Dickson, Claire

Title:
Improving NMR methods to solve 3D molecular structure in solution

Improving NMR methods to solve 3D molecular structure in solution.

Claire Louise Dickson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science.

School of Chemistry, March 2018.

Word Count: 49,100

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is an analytical technique that is used to probe the structure and dynamics of molecules. The relationships between measurable NMR properties and three-dimensional (3D) structure of small molecules were investigated using solution-state NMR spectroscopy with a focus on long-range proton-carbon scalar couplings ($^{n>1}J_{CH}$). The generation and validation of empirical equations to predict three-bond proton-carbon scalar couplings ($^{3}J_{CH}$) from 3D molecular structure were targeted.

The accuracy of selected 2-dimensional methods for measuring $^{n>1}J_{CH}$ in model compounds (strychnine and camphor) was assessed by comparison with 55 $^{n>1}J_{CH}$ measured from coupled ^{13}C spectra. The IPAP-based HSQMBC/HMBC methods offered a balance of experiment time, $^{n>1}J_{CH}$ accuracy and number of measurable $^{n>1}J_{CH}$. However, to maximise the number of $^{n>1}J_{CH}$ measured a combination of multiple experimental methods (coupled ^{13}C, IPAP-HSQMBC and EXSIDE) was required. This led to an average 40% increase in the number of measured couplings compared to using any single technique.

A fragment-based approach to empirical $^{3}J_{CH}$ prediction was developed by using density functional theory (DFT) to calculate $^{3}J_{CH}$ as a function of dihedral angle(s) for over 500 molecules. The fragments consisted of twelve different coupling pathways including saturated/unsaturated centres and β-heteroatoms and a variety of substituent patterns with a focus on the effect of methyl groups.

The fragment-based approach and selected literature empirical equations were validated against experimentally measured and DFT-calculated $^{3}J_{CH}$. This demonstrated that the fragment-based approach could be further improved by the inclusion of a general bond angle correction. Currently, DFT offers the most accurate $^{3}J_{CH}$ calculation with only basis set limitations on the breadth of $^{3}J_{CH}$ that can be calculated. However, the empirical methods offered >2800-fold time saving in the calculation of $^{3}J_{CH}$ compared to DFT when applied to strychnine, camphor and 2-ethyl-1-indanone.

This work is dedicated to my parents,

Sandra and Nigel Dickson.

Acknowledgements

I am sincerely grateful to Professor Craig Butts for his guidance throughout this research project and the process of turning it into a coherent thesis, and for always being so generous with his time.

I would also like to thank Dr Charles Blundell for providing direction and for his comments on this thesis as well as on abstracts, posters and papers over the past four years. I like to think my writing style has improved thanks to this.

I am grateful to Dr Zoltan Takacs for vastly increasing my knowledge of Bruker NMR spectrometers and giving me insight into trials of pulse-sequence programming. I would also like to thank Drs Martin Watson, Misbah Ghafoor and Sadia Mohammed for supporting me during placements at C4X Discovery in Manchester.

To the NMR technicians at the University of Bristol, Mr Paul Lawrence, Ms Rose Silvester and Mr Thomas Leman, I would like to express my gratitude for your technical expertise and patience in teaching me to work with NMR spectrometers. I would also like to thank Thomas Leman for the numerous discussions on handling empirical equations while he was working on his master's research project.

For their extensive help and advice during my time here I thank past and present members of the Butts group, and to Dr Jessica Bame and Dr Ikenna Ndukwe thank you for welcoming me into the group and the world of NMR research in 2014. Any specific contributions of individuals to this thesis are acknowledged as appropriate throughout.

I would also like to thank Dr Natalie Pridmore for sharing her knowledge of thesis writing and X-ray crystallography, Dr Natalie Fey for her guidance in computational chemistry, and Mr Chris Tothill for advice on how best to avoid breaking queuing systems.

To Mr Ben Champion, Ms Esther Townsend, Ms Mary Beton and Ms Varshini Rajkumar, thank you for your kindness, your advice and for listening to my various worries when I thought things were not going to plan.

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (http://www.bris.ac.uk/acrc/) and the Unix computational resources of the School of Chemistry, University of Bristol (http://www.chm.bris.ac.uk/compres/grendel.html).

Finally, I thank C4X Discovery and University of Bristol's EPSRC Doctoral Training Grant (Award 1609475) for funding.

Authors Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Contents

Tables

Figures

Equations

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ADEQUATE	Adequate double quantum transfer experiment
APT	Attached proton test
CASE	Computer-assisted structural elucidation
COSY	Correlation spectroscopy
CPU	Central processing unit
CSD	Cambridge structural database
DEPT	Distortionless enhancement by polarization transfer
DFT	Density functional theory
ES	Expert systems
EXSIDE	Excitation-sculpted indirect-detection experiment
FC	Fermi contact
FID	Free induction decay
GIAO	Gauge-including atomic orbital
GTO	Gaussian-type orbitals
H2BC	Heteronuclear 2-bond correlation
HECADE	Heteronuclear couplings from aSSCI-domain experiments with e.COSY-type cross peaks
HETLOC	Hetero half-filtered total correlation spectroscopy
HF	Hartree-Fock
HMBC	Heteronuclear multiple bond correlation
HSQC	Heteronuclear single quantum correlation
HSQMBC	Heteronuclear single quantum multiple bond correlation
IEFPCM	Integral equation formalism variant of the polarizable continuum model
INADEQUATE	Incredible natural abundance double quantum transfer experiment
INDO	Intermediate neglect of differential overlap
INEPT	Insensitive nuclei enhanced by polarisation transfer
IPAP	In-phase-Anti-phase
MAD	Mean absolute deviation
MCMM	Monte Carlo multiple minimum
MM	Molecular mechanics
MMFF	Merck molecular force field
$^{n}J_{XY}$	Scalar coupling through n bonds between nuclei X and Y
J_{LR}	Long-range scalar coupling constant
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PIP	Pure in phase
PSYCHE	Pure shift yielded by chirp excitation
QM	Quantum mechanics
RMSD	Root mean squared deviation
SD	Standard deviation
TANGO	Testing for adjacent nuclei with a gyration operator
TOCSY	Total correlation spectroscopy
UFF	Uniform force field

Chapter I. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is an analytical technique that is used to probe the structure and dynamics of molecules. By investigating the relationships between measurable NMR properties and molecular structure this work aims to improve methods used to study the three-dimensional (3D) structures of molecules using solution-state NMR spectroscopy.

This introduction (Chapter I), therefore, begins with a brief overview of routinely measured NMR properties (chemical shift and scalar coupling) and their use in two-dimensional (2D) structure determination for establishing the constitution[1] of a molecule, this includes the identity and connectivity of the atoms. The role of molecular modelling in the calculation of NMR properties is then described, including the interpretation of those properties in relation to 3D molecular structure determination, establishing the configuration[2] and conformation[3] of a molecule. The role of empirically predicted scalar coupling constants in 3D molecular structure determination is also discussed.

Chapter II explores the accuracy of experimental measurement of long-range proton-carbon scalar coupling constants ($^{n>1}J_{CH}$) by one-dimensional (1D) coupled ^{13}C spectra and a variety of 2D NMR methods for model compounds, strychnine and camphor. It also includes a comparison of these experimentally measured scalar couplings to values calculated by density functional theory (DFT).

Chapter III and Chapter IV focus on the generation of empirical relationships between three-bond proton-carbon scalar coupling constants ($^3J_{CH}$) and 3D molecular structure. Analogous to the Karplus equation which relates $^3J_{HH}$ to the dihedral angle between the coupled nuclei,[4,5] $^3J_{CH}$ are also dependent on dihedral angle(s)[6,7]; however the literature shows the additional importance of other factors such as the carbon hybridisation state[8], nature of the coupling pathway[9] and substituent electronegativity[7,10,11]. Therefore, empirical relationships with a broader and more general applicability than current literature equations were targeted to account for these factors.

Chapter V and Chapter VI focus on the validation of the new empirical relationships for $^3J_{CH}$ prediction, which was achieved by comparison to the 102 $^3J_{CH}$ measured experimentally using the techniques described in Chapter II. However, since the experimental dataset had a limited structural diversity, two larger datasets of 38,625 and 23,398 $^3J_{CH}$ were calculated using DFT from X-ray crystallographic structures obtained from the Cambridge structural database (CSD)[12].

I.1. Basic principles of NMR spectroscopy

NMR spectroscopy is an important part of the toolbox used by scientists for probing molecular structure. It is a non-destructive technique that is complementary to other spectroscopic and non-spectroscopic techniques available, such as mass spectrometry, infrared spectroscopy, circular dichroism and X-ray crystallography. These techniques each provide different structural information from elemental analysis, to the presence of certain functional groups, to absolute configuration or a complete 3D molecular structure in the case of X-ray crystallography.[13,14] The application of NMR spectroscopy to structural elucidation of solution-state samples is a substantial benefit if obtaining a solid-state crystal/powder is non-trivial for molecule(s) of interest. NMR spectroscopy also provides valuable information about solution-state structure including conformational information, which is particularly important for the study of molecules with biological relevance.[15-19]

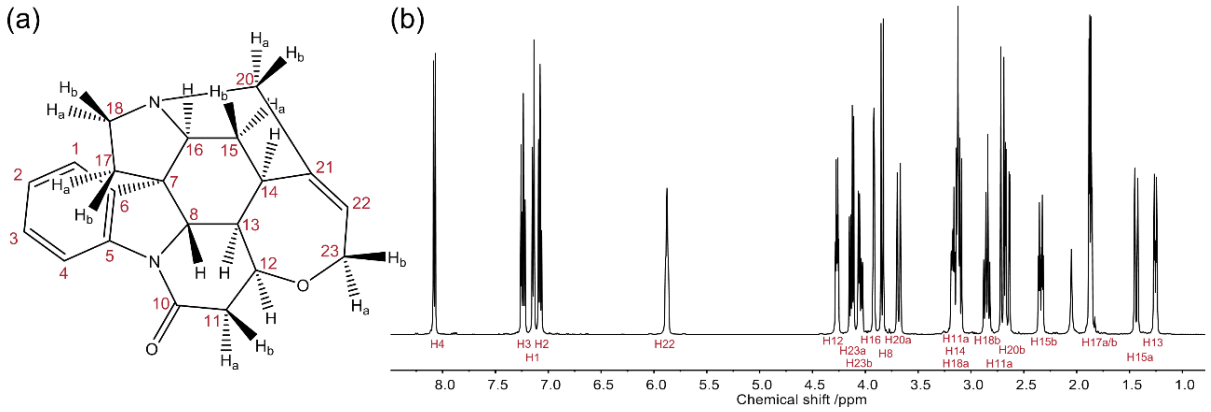


Figure I.1 Labelling system for ^{1}H and ^{13}C in strychnine (a) with a and b identifying prochiral atoms; ^{1}H spectrum of strychnine 500 MHz (b).

Chemists routinely measure the chemical shift of the nuclei in molecule(s) of interest and the scalar couplings between those nuclei by NMR spectroscopy. This information is used, initially, in 2D structural determination, which establishes/confirms the connectivity of the molecule. The 1D ^{1}H spectrum of strychnine (Figure I.1), for example, provides the chemical shifts of the 22 ^{1}H in strychnine ($C_{21}H_{22}O_2N_2$) in addition to long-range proton-proton scalar couplings ($^{n}J_{HH}$) that are measurable by interpreting the multiplicity (splitting) of the peaks. This can be used to confirm the through-bond proximity of atoms; Figure I.2a/b shows the 14.8 Hz $^{2}J_{HH}$ from the geminal protons connected to C20 in strychnine. In 2D NMR spectroscopy, such as ^{1}H-^{1}H correlation spectroscopy (COSY),[20,21] the presence of scalar couplings gives rise to cross-peaks that can be used for structural determination; Figure I.2c shows the corresponding cross-peaks present between H20a and H20b in strychnine.

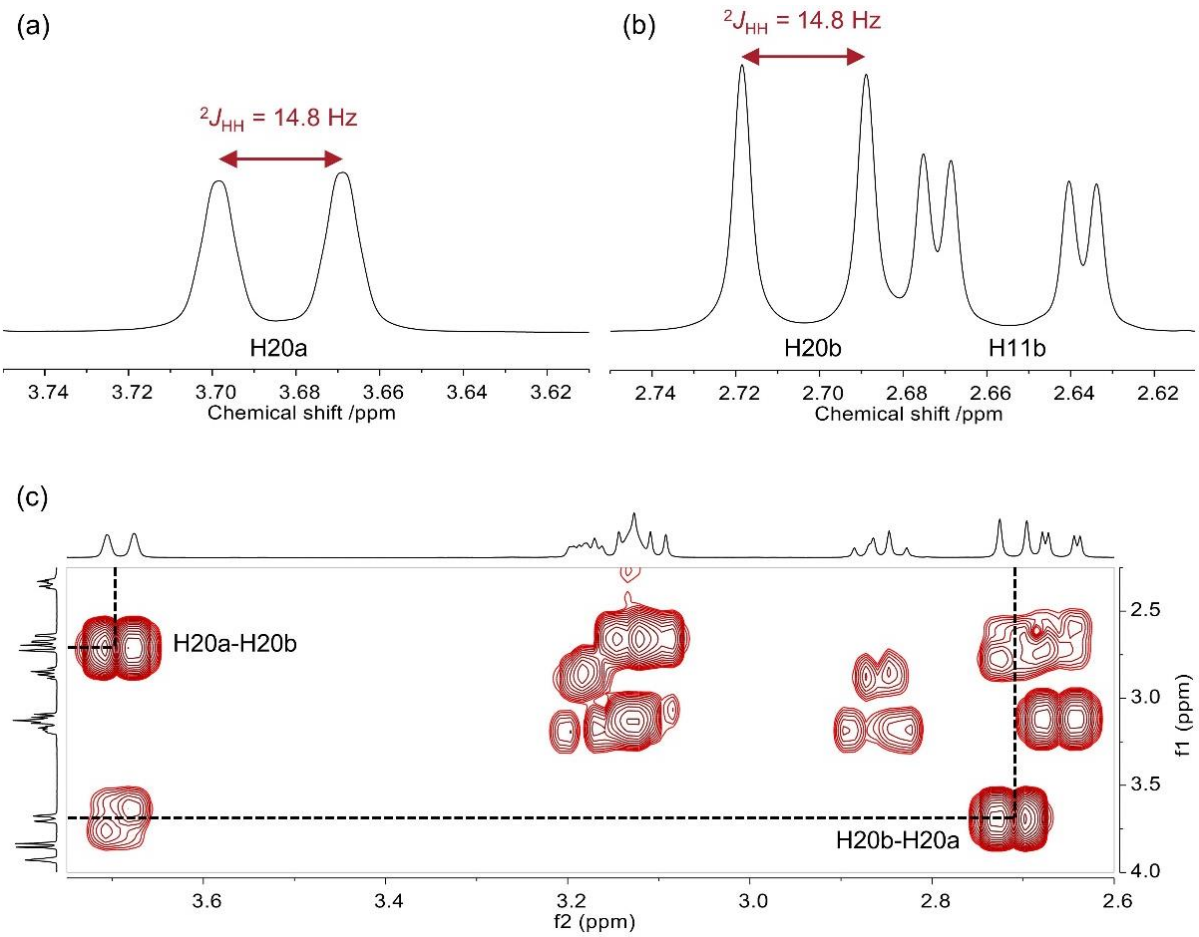


Figure I.2 ^{1}H spectrum of strychnine, H20a (a), H20b and H11b (b) and corresponding section of the ^{1}H-^{1}H COSY spectrum (c) recorded at 500 MHz.

Establishing the connectivity of a molecule is crucial to analysing the 3D structure by NMR because the absence of a robust assignment complicates a conclusive interpretation of further NMR data. NMR provides a comprehensive variety of experimental techniques focussed on determining the connectivity, a selection of which are detailed in Figure I.3, which may be used in cases where the atomic connectivity is not already known from prior data (e.g. X-ray crystallography) or solely for assignment purposes. This list is not exhaustive, being focussed on elucidating the structure of typical organic molecules, that are rich in hydrogen and carbon; additional experiments involving heteronuclei (^{15}N, ^{19}F, ^{31}P or other relatively abundant/enriched NMR-active nuclei) are available depending on the nature of the molecule studied[22].

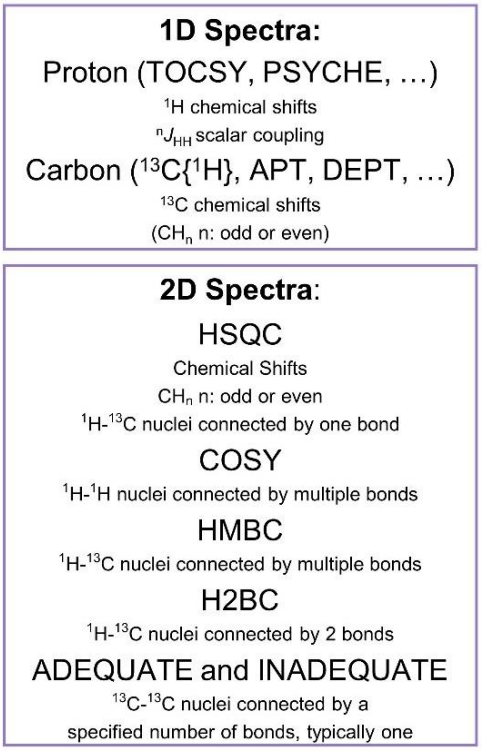


Figure I.3 A selection of common techniques for determining 2D molecular structure with NMR methods. All acronyms defined in Abbreviations section.

The process of structural elucidation for a completely unknown molecule is highly complex and therefore potentially error-prone (particularly for complex natural products); in a review by Nicolaou and Snyder[23] >300 structural revisions were found between 1990 and 2004 (with more than a thousand hits with the SciFinder search parameters used). Maier[24] (in a review of structural revisions covering 2005-2009) notes that "*very often total synthesis is the only way of pinpointing errors in the original structure*" for natural products and grouped the errors in published structures to molecular formula (1D), constitution (2D) and configuration, both relative and absolute (3D).

To reduce potential errors in structural elucidation by NMR a chemist might take one of several approaches: comparing against a model of computed NMR properties[17], expert systems (ES) / computer-assisted structural elucidation (CASE)[28] or total synthesis (for natural products). A recent structural revision[18] of the natural product baulamycin used a synergistic combination of computation of NMR properties and total synthesis.

Computational modelling of experimentally measurable NMR properties (chemical shift, scalar couplings, NOE) start with model(s) of the 3D molecular structure(s). These models may comprise structures with variations in configuration and/or conformation. The chemical shifts and/or scalar couplings that are calculated from the 3D model are compared to experimentally

measured data and the degree of agreement (with a variety of statistical approaches) used to confirm or reject each model[16,17,25,26,27].

The approach of ES/CASE provides an automated framework to generate probable (usually 2D) structures from spectroscopic data with defined logic relating spectral and structural properties. This therefore addresses the 'inverse problem' of structural elucidation, which means that there may not be a unique solution, depending on the information provided.[28] Therefore, the set of rules relating spectral and structural properties generally include an element that compares the experimental spectral data to predicted spectral data from the generated structures, which may be by empirical means (such as chemical shift libraries or Karplus equations) or by molecular modelling (such as DFT chemical shift calculations). The automated nature of CASE means that, given the right workflow, a large amount of spectral data can be processed, while also saving time relative to performing the same analysis 'by-hand', with the benefit of probabilistic ranking of the structures with a replicable workflow.[29]

Solving 3D molecular structure is more challenging, two examples of CASE systems for this purpose are the ADC/Structural Elucidator (StrucEluc)[30] and CASE-3D[31,32]. StrucEluc augments chemical shift and connectivity data (from 1D and 2D NMR spectra) with NOESY/ROSEY data to determine relative stereochemistry[30] and CASE-3D uses NMR-data from both isotropic samples (solution state) and anisotropic sample (aligned media), combining chemical shift, scalar coupling and NOESY data with residual dipolar coupling (RDC) and residual chemical shift anisotropy (RCSA) data to determine relative stereochemistry and preferred conformation of molecules.[31,32] Other approaches, such as the NMR methodology of Blundell *et al.*,[15] use a dynamic model of 3D molecular structure to focus on using experimentally measured NOE, scalar coupling and RDC data to determine preferred conformations.

Therefore, any additional information or improvement in accuracy can make a significant contribution to the model and the quality of the determined structure. In order to improve methods for the determination of structure it is first necessary to understand in more detail the origins of measurable NMR properties, chemical shift (section I.1.1) and scalar coupling (section I.1.2) as well as the current progress in methods that aid the interpretation with respect to structure, in particular quantum mechanics based molecular modelling (section I.2) and empirical predication of scalar couplings (section I.3).

I.1.1. Chemical shift

In the Standard Model, elementary particles have certain intrinsic quantum properties including mass, spin and charge; combinations of these particles form molecules and the nuclei within them. As composite particles, nuclei possess a related property known as nuclear spin angular

momentum which is characterised by a nuclear spin quantum number (I) which takes integer or half integer values (=0, ½, 1, 1½ …). Nuclei (and electrons) additionally possess an intrinsic magnetic moment (***μ***) which is related to the spin angular moment (***S***) as follows (Equation I.1)[33,34]:

$$\boldsymbol{\mu} = \gamma \boldsymbol{S} \qquad \text{Equation I.1}$$

Where γ is the magnetogyric (gyromagnetic) ratio.[33]

The magnetogyric ratio varies for different nuclei, ($\gamma(^1H) \approx 267.522\times10^6$ rad s^{-1} T^{-1}, $\gamma(^{13}C) \approx 67.283\times10^6$ rad s^{-1} T^{-1})[33] and the sign of the value determines whether the magnetic moment is parallel or anti-parallel to the spin angular momentum. In the presence of a static magnetic field the spin of the nucleus precesses about the external static magnetic field, B_0, therefore the magnetic moment traces a cone (Figure I.4) about B_0.

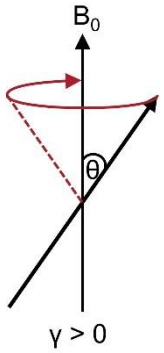


Figure I.4 The precession of the magnetic moment of a nucleus in a static magnetic field B_0, where γ determines the direction of motion.[33]

The frequency of the precession, known as the Larmor frequency, is related to the strength of the magnetic field as follows (Equation I.2):

$$\omega_0 = -\gamma B_0 \text{ (in rad } s^{-1}\text{)} \qquad \text{Equation I.2}$$

This equation also provides a basis for understanding the distinction between different nuclei within molecules. The frequency that a nucleus precesses at, ω_0, is related to the external magnetic field by the magnetogyric ratio of that nucleus. Therefore, different elements and isotopes can be differentiated; for example, in an 11.7 T external field ^{1}H resonate at 500 MHz and ^{13}C at approximately 125 MHz.

However, it can be seen in the ^{1}H spectrum of strychnine in Figure I.1, the nature of the nucleus observed, and the strength of the external magnetic field are not the only factors that determine the resonant frequency, it is also affected by local effects within the sample.

$$B_{local} = B_0 - \sigma B_0 \qquad \text{Equation I.3}$$

Relating the effective local field strength, B_{local}, experienced by a nucleus to the external magnetic field strength, B_0. Where σ is a (dimensionless) shielding constant.[35]

The shielding constant (σ) in Equation I.3 is formed by the sum of two local contributions, a diamagnetic (σ_{dia}) term, a paramagnetic (σ_{para}) term, and a third term, σ', originating from the effects of neighbouring groups and other intermolecular interactions such as those arising from solvents (Equation I.4). These contributions to σ may augment or oppose the external magnetic field, deshielding or shielding the nucleus respectively.[35]

$$\sigma = \sigma_{dia} + \sigma_{para} + \sigma' \qquad \text{Equation I.4}$$

The shielding constant σ is a sum of three contributions.[35]

Diamagnetic effects oppose the external field and decrease the effective field. For nuclei like ^{1}H this lower effective field at the nucleus arises as the external field, B_0, induces circulation of the electrons in the atomic/molecular orbitals. This movement of charged particles will therefore produce a magnetic field that opposes B_0 and results in the nucleus appearing shielded from the magnetic field as it is experiencing a lower effective field, B_{local}. The size of this effect depends on the electron density close to the nucleus in the ground state electronic wavefunction. Diamagnetic contributions to the shielding constant are responsible for the intuitive effects of electron withdrawing and electron donating groups, particularly in ^{1}H spectra. This effect illustrates that the chemical environment affects the frequency that a nucleus resonates at, therefore chemical shifts can be useful for chemical structure determination.[36]

Paramagnetic contributions to σ also arise from the circulation of electrons, however in this case the magnetic field generated augments the external magnetic field. This effect arises through the mixing of excited and ground state electronic wavefunctions due to the application of the external field.[37] This results in new pathways for electrons to circulate; the spherical symmetry of s-orbitals averages this term to zero[37] therefore this effect is more prominent in spectra of nuclei with *p*-orbitals such as ^{19}F NMR rather than ^{1}H NMR.[38]

The effects that comprise σ' may be diamagnetic or paramagnetic in origin and arise from a wide variety of inter- and intramolecular sources.[36] Neighbouring group effects include those from nearby π systems such as aromatic rings, alkenes or alkynes. Hydrogen-bonding within the molecule or intermolecular sources (such as solvent) generally has a deshielding effect on the ^{1}H in the bond.[39] The careful introduction of σ' effects[40] with chiral solvating agents or chiral derivatizing agents such as the Mosher ester analysis[41] can aid in the determination of absolute configuration.

It is important to note that rather than reporting the frequency in Hertz, chemical shifts are reported in parts per million (ppm) compared to a reference compound (Equation I.5). Calculated this way, these values become independent from the magnetic field strength of the NMR spectrometer, thereby facilitating comparison of spectra between different instruments.

$$\delta(\mathrm{ppm}) = 10^6 \times \frac{\nu - \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}}$$ Equation I.5

Calculation of chemical shift, where ν is the frequency of the peak of interest, ν_{ref} is the frequency of a reference compound and δ is the resulting chemical shift in ppm.[42]

I.1.2. Scalar coupling

The ^{1}H spectrum of strychnine (Figure I.1, enlarged sections in Figure I.2a/b) shows that the signals from the different ^{1}H nuclei in the molecule can have complex fine structure rather than appearing as single peaks (singlets). The splitting of the signal is caused by interactions between the spins, which are mediated by chemical bonds. This effect is referred to as indirect spin-spin coupling, *J*-coupling, or scalar coupling (in isotropic solution, which is the focus of this work) and the magnitude of the splitting is set by the coupling constant (*J*). Figure I.5 summarises the results of quantum mechanical treatment of scalar coupling in terms of an energy level diagram[35], a more in-depth description of the mechanism of scalar coupling is returned to in section I.2.

The simplest example is scalar coupling between two nuclei A (I = ½) and X (I = ½), in the absence of any coupling the spin of nucleus A has no influence on the spin of nucleus X and vice versa. This is seen in the energy level diagram in Figure I.5a, where the allowed (single quantum) transitions give rise to the A (light red arrows) and X (dark red arrows) singlets. However, when coupling is present between A and X, certain combinations of spins are stabilised by the interaction while others are destabilised (Figure I.5b) depending on the sign of the coupling. Therefore, the transitions between energy levels ($\nu_A \pm \frac{1}{2}J$ and $\nu_X \pm \frac{1}{2}J$) are no longer equivalent resulting in a doublet for each signal.

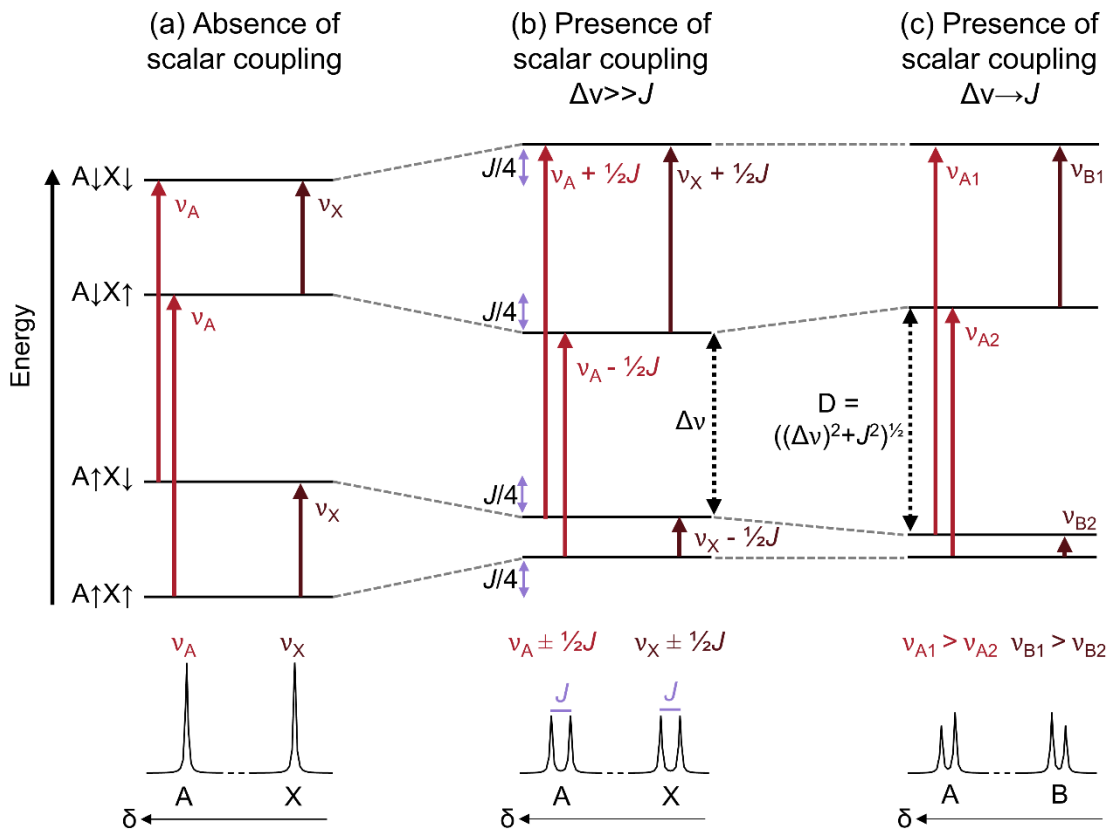


Figure I.5 Energy level diagram for two spins which are not coupled (a), are coupled (b), or are strongly coupled (c), the energy differences are exaggerated for clarity. Diagram based on references [37,36].

The appearance of a peak may be further complicated by the effects of strong coupling (also known as second order effects) which can distort splitting patterns, relative peak heights and lineshapes away from those predicted simply using Figure I.5b. Continuing with the simplest case of two nuclei, A (I = ½) and B (I = ½), where (unlike nuclei A and X), A and B are close in chemical shift and when coupled, a simple doublet with equal intensity in each half is no longer observed (Figure I.5c). The lower intensity of the outer peaks relative to the inner peaks of the pair is known as 'roofing'. As the chemical shifts of A and B become closer ($\Delta\nu \rightarrow 0$) the degree of roofing within the peaks increases (Figure I.6a-d) to the point where $\Delta\nu = 0$ and both A and B appear as a single peak (Figure I.6e).

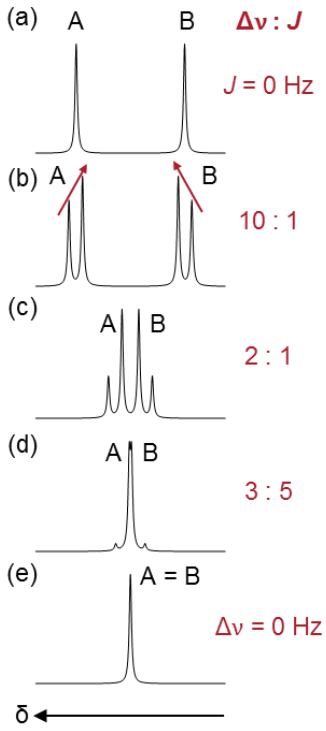


Figure I.6 Simulated spectra of two nuclei, A and B, without scalar coupling (a) and with scalar coupling as Δν→0 (b) to (e).

Figure I.5c summarises the results of a quantum mechanical treatment of two coupled spins in terms of an energy level diagram, with the transition energies and relative intensities are in Table I.1.[35] As the difference (Δν) between the two central energy levels with mixed spin states becomes small the corresponding wavefunctions can mix. This stabilises one energy level (A↑B↓) and destabilises the other (A↓B↑). This mixing affects which transitions are allowed, and as the chemical shifts of A and B converge the transitions that correspond to the outer peaks have a reduced probability and therefore decrease in intensity until they no longer appear when Δν = 0 (Figure I.6e). These effects are not limited to pairs of nuclei and may involve multiple nuclei, which can lead to highly complex multiplets.[43]

Table I.1 Calculated transition energies and relative intensities for a coupled AB spin system.[35]

Transition label	Transition energy	Relative intensity
ν_{A1}	$\frac{1}{2}(\nu_A + \nu_B) + \frac{1}{2}J + C$	$1 - \sin 2\theta$
ν_{A2}	$\frac{1}{2}(\nu_A + \nu_B) - \frac{1}{2}J + C$	$1 + \sin 2\theta$
ν_{B1}	$\frac{1}{2}(\nu_A + \nu_B) + \frac{1}{2}J - C$	$1 + \sin 2\theta$
ν_{B2}	$\frac{1}{2}(\nu_A + \nu_B) - \frac{1}{2}J - C$	$1 - \sin 2\theta$

The angle 2θ is defined as $\Delta\nu/2C = \cos 2\theta$ and $J/2C = \sin 2\theta$, where $C = \frac{1}{2}D$.

I.2. Quantum mechanical calculation of NMR parameters

Establishing a computational model of experimentally measurable NMR properties (chemical shift, scalar couplings, NOE) of a molecule first requires a model of the 3D molecular structure.

In solution, flexible molecules interconvert between different conformations due to rotation about single bonds (depending on the definition used this may also include changes in geometry due to inversion[3]). The structural variations in the conformers may correspond to variation in the NMR properties. Therefore, assuming the interconversion between conformers is rapid compared to the NMR timescale, the observed NMR properties will reflect a population weighted average (according to the Boltzmann distribution) of all the different conformers present in solution.

Therefore (as outlined in Figure I.7), when generating a computational model of a molecule it is necessary to obtain the different conformers that are populated in solution for accurate comparison to experimental data. This is achieved through conformational searching, which can be performed with several different methods including systematic search algorithms,[44,45] model-building methods,[46-48] random approaches,[49-52] distance geometry[53] and molecular dynamics methods.[55]

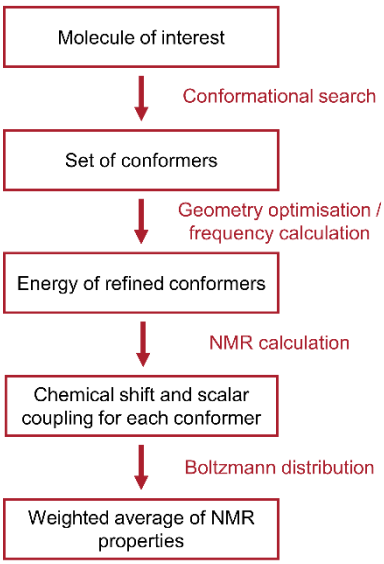


Figure I.7 Workflow to generate a computational model of solution state NMR properties.

I.2.1. Density functional theory (DFT)

In this thesis, density functional theory (DFT) is used for all quantum mechanics (QM) calculations including geometry optimisations, frequency calculations and NMR calculations. The foundation of DFT stems from the theorem of Hohenburg and Kohn[54] which states that the ground state electron density determines the total electronic energy of a molecule:

$$E[\rho(\boldsymbol{r})] = \int V_{ext}(\boldsymbol{r})\rho(\boldsymbol{r})d\boldsymbol{r} + F[\rho(\boldsymbol{r})] \qquad \text{Equation I.6}$$

The ground state energy, E, of a system is a functional of the electron density $\rho(\mathbf{r})$, $V_{ext}(\mathbf{r})$ is an external potential (such as the Coulomb interaction with the nuclei) and F is a functional which gives the sum of kinetic and interelectronic energies[55,56].

The Kohn-Sham formulation[57] treats the electron density as the sum of one-electron orthonormal orbitals as follows:

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2 \qquad \text{Equation I.7}$$

Form of the electron density $\rho(\boldsymbol{r})$, where $\psi_i(\boldsymbol{r})$ is the i^{th} one-electron Kohn-Sham orbital and N is the total number of electrons[55,56].

I.2.1.1. Basis Sets

In DFT $\psi_i(\boldsymbol{r})$, the one-electron Kohn-Sham orbital (Equation I.7) can be formed from a linear combination of atomic orbitals:

$$\psi_i(\boldsymbol{r}) = \sum_{k=1}^{M} c_{ki}\varphi_k \qquad \text{Equation I.8}$$

Construction of ψ_i, using atomic orbitals $\boldsymbol{\phi}_k$ where c_{ki} are constants, determined to give the lowest energy wavefunction for the M atomic orbitals in the basis set.[58]

These basis sets are constructed of separate (primitive) functions, generally grouped to form core and valence orbitals, and compatible with additional modifications to improve the description of molecular orbitals with polarisation[59] or diffuse functions[60] (with associated computational costs). A minimal basis set uses one basis function per atomic orbital, while a split-valence basis uses two or more basis functions for the valence atomic orbitals. The Pople basis sets are split-valence basis sets, that use Gaussian (primitive) functions to form the basis functions such as 3-21G,[61] 6-31G,[62-68] 6-311G(d,p)[69,70]. The numbers which describe the Pople basis sets "*K*-*LMN*G" indicate the number of primitive Gaussian functions which form each basis function: *K* the number of core Gaussian functions and *L*, *M*, and *N* the valence Gaussian functions for each of the split-valence basis functions. The computational efficiency of basis sets is increased by forming the basis function from a linear combination of the primitive functions and is further improved by reducing the number of optimised parameters by sharing the value of the Gaussian exponents between the different basis functions; this process is known as contracting.[71]

The accurate calculation of NMR properties such as scalar couplings can be crucially dependent on the description of electron density at the nucleus, something that is not always well described in standard (valence oriented) basis sets. Therefore, improvements to DFT methods to calculate scalar couplings have typically focussed on introducing additional core/tight functions and optimal contractions with respect to the core orbitals for traditionally valence-oriented basis sets, such as Jensen's polarisation consistent pc*J*-n basis sets,[72-74] the

modified Pople basis sets (6-31G-*J* and 6-311G-*J*)[75] and the correlation consistent aug-cc-pVTZ-*J* basis sets[76,77].

Another approach is the "mixed" method implemented in Gaussian09[78], which uses an uncontracted basis set with additional tight *s* functions to calculate the Fermi contact (FC) contribution to the coupling, and then the contracted (unmodified) basis set is used to calculate the spin dipolar, paramagnetic spin-orbit and diamagnetic spin-orbit contributions to the coupling. This process improves the basis set description of electron density at the nucleus for valence-oriented basis sets for the FC term and therefore improves the accuracy of calculated scalar couplings (tested against the uncontracted UGBS2P basis set).[79]

I.2.1.2. Functionals

Kohn-Sham[57] give the $F[\rho(\boldsymbol{r})]$ term in Equation I.6 as the sum of E_{KE}, the kinetic energy for a non-interacting system, E_H, the classic Hartree electrostatic energy and E_{XC}, the exchange-correlation energy.[55]

$$F[\rho(\boldsymbol{r})] = E_{KE}[\rho(\boldsymbol{r})] + E_H[\rho(\boldsymbol{r})] + E_{XC}[\rho(\boldsymbol{r})] \qquad \text{Equation I.9}$$

The exact form of the exchange-correlation energy is not known. However, it is possible to make approximations to obtain forms for the exchange-correlation functional. Three different examples of functionals (B3LYP,[80,81] mPW1PW91[82] and ωB97X-D[83,84]) and their practical benefits with respect to the calculation of NMR properties are explored within the results of this thesis.

I.2.1.1. Inclusion of solvent effects

In this thesis, the DFT compatible Integral Equation Formalism (IEF) variant of the Polarizable Continuum Model (PCM) has been used when solvent effects are included in calculations. The simplest form of PCM considers a homogenous, isotropic solvent which has only electrostatic interactions with the solute.[85] By assuming the solution is infinitely dilute only solute-solvent effects need be considered and therefore the solute molecule can be considered within a cavity surrounded by a constant dielectric medium representing the solvent. The IEFPCM model does not account for all solvent effects; as an example, hydrogen bonding is not considered.

I.2.2. Calculation of NMR properties

In the 1950s Ramsey published a series of papers[86-91] on the calculation of chemical shift and spin-spin coupling constants.

In this thesis, the Gauge-including atomic orbitals (GIAO) method has been used to calculate NMR properties using DFT.[92,93] As previously mentioned (section I.1.1) the chemical shift of a nucleus depends on the shielding of the nucleus from the external magnetic field by the surrounding electrons in the molecule. This shielding depends on the relative orientation of

external magnetic field and the electron density in the molecule, Equation I.10.[94] The isotropic component of σ, can then be calibrated relative to an external or internal reference nucleus to obtain chemical shifts.

$$\sigma_{ij}^{N} = \frac{\partial^2 E}{\partial \boldsymbol{\mu}_i^N \partial \boldsymbol{B}_j} \qquad \text{Equation I.10}$$

The magnetic shielding tensor. Definitions: σ, magnetic shielding tensor; E, total electronic energy; $\boldsymbol{B}$ the external magnetic field; $\boldsymbol{\mu^N}$ the magnetic moment of nucleus N; i, component of the external magnetic field; j, component of the induced magnetic moment.[95]

The works by Ramsey and Purcell[95] and Ramsey[90] calculated spin-spin couplings as formed of four terms: Fermi Contact (FC), diamagnetic spin-orbit (DSO), paramagnetic spin-orbit (PSO) and spin-dipole (SD). The FC term generally has a dominant contribution to scalar couplings, (and to reduce the computational costs may be the only term considered), however all four mechanisms should be considered for accurate calculation of scalar coupling and have been adapted for compatibility with DFT treatment of the electronic states of a molecule.[96-99]

I.2.2.1. Mechanism of Scalar Coupling

Koide and Duval[100] combined valence bond theory with the Ramsey formula for the FC term (Dirac vector model) to correctly predict the sign of $^{n}J_{HH}$ in saturated compounds for different numbers of intervening bonds (n = odd or even).

$$J_{NN'} \propto -\gamma_N \gamma_{N'} \langle 0 | \boldsymbol{s_N} \cdot \boldsymbol{s_{N'}} | 0 \rangle \qquad \text{Equation I.11}$$

Where $J_{NN'}$ is the coupling between nuclei N and N', for electron spins $\mathbf{s_N}$ and $\mathbf{s_{N'}}$.

Scalar coupling does not arise from the direct dipolar interaction of the spins of two nuclei, because this anisotropic interaction averages to zero due to molecular tumbling in solution.[33] Instead, scalar coupling is mediated through chemical bonds. Like nuclei, electrons possess spin (S = ½) and have a magnetogyric ratio which is many times greater than ^{1}H in magnitude. Unlike the dipolar interactions between two nuclei the interactions between electrons and nuclei do not average to zero in solution if the separation between the electron and the nucleus is comparable to the nuclear radius. This interaction is determined by the isotropic Fermi Contact interaction (Equation I.12). The probability of finding an electron at the nucleus is related to the magnitude of this interaction, and therefore requires electron density (s character) at the nucleus in the electronic wavefunction.[34]

$$contact\ interaction \propto -\gamma_e \gamma_n \boldsymbol{I} \cdot \boldsymbol{S} \qquad \text{Equation I.12}$$

Fermi Contact interaction between an electron, e, with magnetogyric ratio γ_e and spin angular momentum **S** and a nucleus, n, with magnetogyric ratio γ_n and spin angular momentum **I**.[37]

Equation I.12 also indicates that antiparallel arrangements of electron and nuclear spins are stabilised, and parallel arrangements are destabilised for nuclei that have a positive magnetogyric ratio, such as ^{1}H or ^{13}C. This interaction provides the first step of the mechanism through which the spins of nuclei can couple (and which Koide and Duval[100] used to predict the sign of $^{n}J_{HH}$). Consider the model of the $^{12}CH_2$ group:

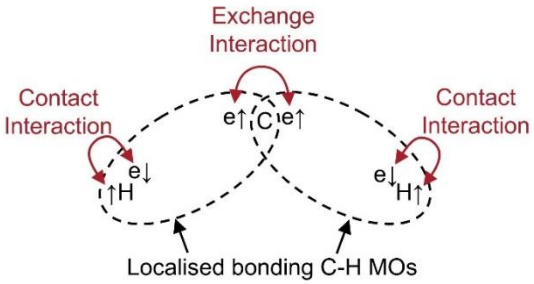


Figure I.8 Example pathway by which the two ^{1}H nuclei in a $^{12}CH_2$ group can couple. Diagram based on references [34,37].

The Fermi contact interaction between the spin of the ^{1}H nucleus and one of the electrons in the C-H bond polarises the bond. This means that there is a higher probability of finding the electron that is in an antiparallel arrangement (Pauli principle[101]) close to the ^{1}H nucleus and electron with a parallel arrangement further from the nucleus. The exchange interaction (Hund's rule[102]) between electrons which favours a parallel arrangement links the two C-H molecular orbitals (MO). The example in Figure I.8 shows a simple model pathway through which electrons can mediate the coupling of nuclei, wherein the two ^{1}H nuclei are stabilised in configuration of H↑H↑ (a negative scalar coupling from Ramsey's FC term). For an odd number of intervening bonds ($^{1}J_{CH}$ or $^{3}J_{HH}$) the coupling is positive and the H↑H/C↓ configuration stabilised, such as the example in Figure I.5(b).

I.3. Scalar coupling and 3D structure

This section describes the relationship of scalar couplings to the 3D structure of a molecule in further detail. Throughout this thesis the notation $^{n}J_{XY}$ is used to refer to scalar coupling between nuclei X and Y through n bonds. To indicate intermediate nuclei along the coupling pathway $^{n}J_{XABY}$ is used where X and Y are coupled with nuclei A and B between them.

I.3.1. Long-range proton-proton scalar couplings: $^{n}J_{HH}$

The Karplus equation (Equation I.13) relates three-bond proton-proton scalar couplings ($^{3}J_{HH}$) to the dihedral angle (Φ) between the coupled nuclei (Figure I.9). It was developed as an empirical approximation of the results of calculating the contact electron-spin contribution (FC term) to $^{3}J_{HCCH}$ using valence-bond theory for an HCCH fragment with Φ varied from 0° to 180° in 30° steps.[4,5]

$$^{3}J_{HH} = A + B\cos\Phi + C\cos 2\Phi$$ Equation I.13

Where A, B and C are constants and A = 4.22, B = -0.50 and C = 4.50 Hz for sp^3 hybridised carbons with a C-C bond length of 1.543Å.[5]

The term "Karplus equation" is often used interchangeably to mean the general form of Equation I.13, the exact values of the coefficients (A, B and C) determined by Karplus or less frequently any empirical equation which relates scalar coupling to dihedral angle(s). In this thesis the "Karplus equation" is used to refer to the general form of Equation I.13, with coefficients that lead to $^{3}J_{XY}$ minima at Φ = ±90°, local maxima at Φ = 0° and global maxima at Φ = ±180°.

Although the dihedral angle (Φ) between the coupled nuclei (Figure I.9) is the dominant geometric parameter in literature equations predicting $^{3}J_{HH}$, Karplus noted in 1963 that there are many other structural factors such as hybridisation (of the coupling pathway), bond length, bond angle and substituent electronegativity that have significant effects on the magnitude of the coupling.[5]

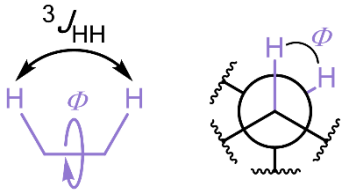


Figure I.9 $^{3}J_{HH}$ coupling pathway, defining the dihedral angle Φ.

Attempts to improve the accuracy of the empirical relationships between $^{3}J_{HH}$ and 3D molecular structure can be approximately grouped into two approaches. The first is to generate multiple sets of coefficients for the Karplus equation by targeting related fragments within specific sets of compounds such as nucleic acids,[103] peptides/proteins[104-106] and carbohydrates[107]. A drawback of this method is the need for $^{3}J_{HH}$ datasets of similar compounds with large enough variation in the dihedral angle to describe the full Karplus curve. The second method is to modify the Karplus equation by introducing additional variables to produce a general relationship between the coupling constant and the substituent, therefore the dataset need not be limited to only related compounds.

The equation developed by Haasnoot, de Leeuw and Altona, for example, introduced the substituent electronegativity and relative orientation to the Karplus equation (rearranged in the first three terms of Equation I.14)[108]:

$$^3J_{HH} = P_1 \cos^2 \Phi + P_2 \cos \Phi + P_3 + \sum \Delta\chi_i (P_4 + P_5 \cos^2(\xi_i \Phi + P_6 |\Delta\chi_i|)) \quad \text{Equation I.14}$$

Empirical $^3J_{HH}$ equation accounting for substituent effects. $P_{1\text{-}6}$ are constants specific to the number of substituents, $\Delta\chi_i$ is the electronegativity of substituent i relative to hydrogen on the Huggins scale[109] and ξ_i is 1 or -1 depending on the orientation of substituent *i* as defined in Figure I.10.

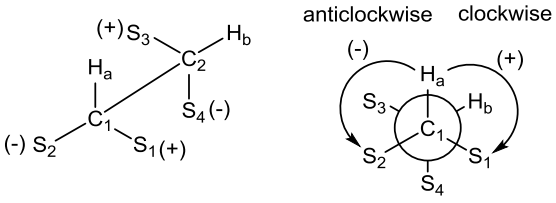


Figure I.10 Definition of positive and negative substituents in Equation I.14.

Equation I.14 was parametrised with a dataset of 315 experimentally measured coupling constants from 109 conformationally rigid compounds that were modelled computationally to obtain the relevant dihedral angles. This obtained five sets of parameters giving one general equation, one equation which included a correction for the effects of β-substituents (using Equation I.15) and three equations specific to the number of substituents. For $^3J_{HH}$ α-substituents refer to substituents directly bonded to the coupling pathway and β-substituents those two bonds from the coupling pathway. When comparing predicted $^3J_{HH}$ to the experimental dataset, Equation I.14 offered a statistical improvement over the standard Karplus equation that had been parameterised with the same dataset.

$$\Delta\chi^{group} = \Delta\chi^{\alpha} - P_7 \Sigma \Delta\chi^{\beta} \quad \text{Equation I.15}$$

Correction to relative electronegativity to include β-substituents.[108]

Ōsawa *et al*.[110,111] further extended the approach of Haasnoot *et al*. to include bond angle and bond length corrections for $^3J_{HH}$. This also used a set of experimentally measured scalar couplings and geometries calculated by molecular mechanics to parameterise a general equation for $^3J_{HH}$ that included dihedral angle (θ in Equation I.16) and α- and β-substituent electronegativity effects:

$$
\begin{aligned}
{}^{3}J_{\mathrm{HH}} = {} & \mathrm{M} + \mathrm{A}\cos\theta + \mathrm{B}\cos 2\theta + \mathrm{C}\cos 3\theta + \mathrm{D}\cos^{2} 2\theta \\
& + \mathrm{W}\left(\mathrm{E}\cos\theta \sum \Delta\chi_i \cos\varphi_i + \mathrm{F}\sum \Delta\chi_i \cos 2\varphi_i \right. \\
& \left. + \mathrm{G}\sum \Delta\chi_i\right) + \mathrm{H}\left(\frac{\omega_1 + \omega_2}{2} - 110\right) \\
& + \mathrm{I}(r_{\mathrm{cc}} - 1.5) + \mathrm{K}\sum \Delta\chi_j^{\beta} \cos 2\psi_j + \mathrm{L}r^{-4}
\end{aligned}
\qquad \text{Equation I.16}
$$

Where A-M are constants, θ is the dihedral angle between the coupled ^{1}H, i is the number of α-substituents, j is the number of β-substituents, $\Delta\chi_i$ is the difference in electronegativity to H for α-substituent i, $\Delta\chi_i^{\beta}$ for β-substituent j, φ_i is the dihedral angle between α-substituent i and the coupled ^{1}H, ψ_j between β-substituent j and the coupled ^{1}H, ω_1 and ω_2 the HCC bond angles, r_{CC} the carbon-carbon bond length, and r the distance of the coupled ^{1}H to non-bonded C/O <3.3 Å.[110]

The relationship between the value of $^{3}J_{HH}$ and the dihedral angle between ^{1}H nuclei gives a clear link to 3D molecular structure, although the accuracy of empirically predicted $^{3}J_{HH}$ values may vary due to the other effects discussed and the prediction may be improved by their inclusion.

I.3.1.1. Two-bond proton-proton scalar couplings: $^{2}J_{HH}$

Substituents, hybridisation (of the coupling pathway), bond angle and bond length also affect the magnitude (and sign) of $^{2}J_{HH}$.[112] For example, the direct inductive effects of substituents attached to carbon typically lead to an increase in $^{2}J_{HCH}$ where the substituent is electron withdrawing and a decrease where the group is electron donating.[112] However, substituents that are bonded to the carbon may have additional effects that can be used probe the 3D structure of a molecule. The effect of lone pairs or adjacent π systems of substituents on $^{2}J_{HH}$ depends on the orbital overlap with the C-H bonds and therefore $^{2}J_{HH}$ can indicate the orientation of the C-H bond relative to these groups.

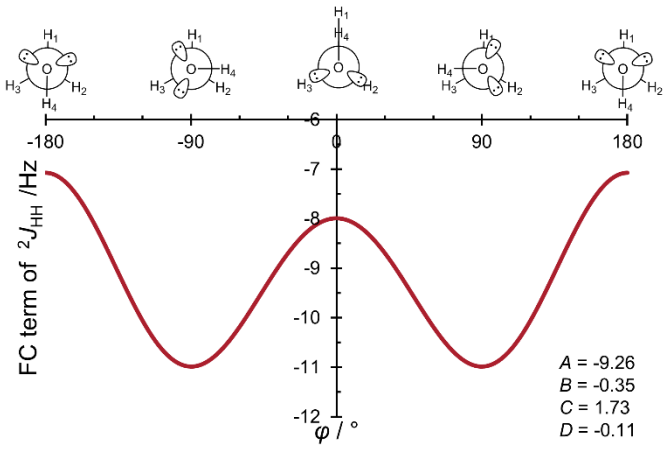


Figure I.11 The relationship of $^{2}J_{HH}$ for H_2 and H_3 to the lone pairs on an adjacent substituent (Equation I.17). The dihedral angle φ is defined as between H_1 on the methyl group and H_4 on the alcohol. The calculations to obtain the FC term were performed using FPT (Finite Perturbation Approach) DFT with the B3LYP functional and GAMESS-pVTZ basis set.[112]

Figure I.11 illustrates the effect of rotation of the HOCH dihedral angle (φ) in methanol on the FC term of $^{2}J_{HH}$ as described by Equation I.17[112]. In this example, the magnitude of the $^{2}J_{HH}$ is lowest when the oxygen lone pairs are aligned such that donation into the H-C σ anti-bonding orbitals is possible.

$$^{2}J_{\mathrm{HH}} = \mathrm{A} + \mathrm{B}\,cos\,\varphi + \mathrm{C}\,cos\,2\varphi + \mathrm{D}\,cos\,3\varphi \qquad \text{Equation I.17}$$

I.3.2. Proton-carbon scalar couplings: $^{n}J_{CH}$

Proton-carbon scalar coupling constants, $^{n}J_{CH}$, are a powerful tool in the elucidation of 3D structure, in particular for parts of molecules that have few ^{1}H and therefore provide insufficient $^{n}J_{HH}$ or ^{1}H-^{1}H NOESY data for determining their 3D structure.

I.3.2.1. One-bond proton-carbon scalar coupling: $^{1}J_{CH}$

The hybridisation of carbon is a major contributing factor to the size of the one-bond coupling constant between ^{1}H and ^{13}C (Table I.2). $^{1}J_{CH}$ increases with the increasing *s*-character of the coupled carbon,[113,114] however other effects such as bond angles, bond lengths and substituent effects also impact $^{1}J_{CH}$ [8]. For example, the addition of electron withdrawing groups (X, Table I.2) to the ^{13}C typically leads to an increase in the magnitude of the coupling constant. However, as for $^{n}J_{HH}$ the effects of substituents are more complex than the sole impact of electronegativity.

Table I.2 Typical ranges for $^{1}J_{CH}$ depending on carbon environment.[22]

Carbon Environment	$^{1}J_{CH}$ /Hz
Alkane	125-135
Alkane with C-X (X = N, O, S)	135-155
Alkene	155-170
Alkyne	240-250
Aromatic (benzene)	155-165

The impact of lone pairs (for example) and other stereoelectronic effects on $^{1}J_{CH}$ are known as Perlin effects.[115,116] Generally, where the lone pair(s) of a substituent are aligned such that there is an interaction with the CH orbitals there is a decrease in the size of $^{1}J_{CH}$. In the absence of other effects, *trans* arrangements of the CH bond and lone pair correspond to smaller $^{1}J_{CH}$ than *cis* arrangements,[112] Figure I.12 gives an example of this effect in 5-*t*-butyl-1,3-dioxane.

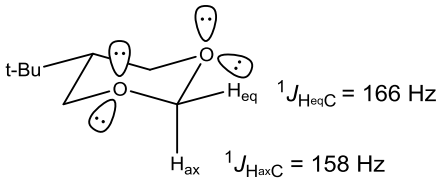


Figure I.12 Experimentally measured $^{1}J_{CH}$ for the compound 5-t-butyl-1,3-dioxane.[117]

I.3.2.2. Two-bond proton-carbon scalar coupling: $^{2}J_{CH}$

Generally, $^{2}J_{HC}$ are small in magnitude and may be positive or negative in sign, as summarised in Table I.3. The factors that affect $^{2}J_{CH}$ are analogous to those which affect $^{2}J_{HH}$ such as the bond angles and lengths between the two coupled nuclei or the hybridisation of the coupling pathway and any attached substituents.[8] However, (compared to $^{2}J_{HH}$) additional effects arise from variation in the coupled ^{13}C (the α-carbon) such as hybridisation state and additional α-substituent effects. The coupling pathway for proton-carbon scalar couplings uses labels referring to the coupled ^{13}C as the α-carbon, with the intermediate coupling pathway similarly labelled. For example, the coupling pathway is H-Cβ-Cα for $^{2}J_{CCH}$ and for $^{3}J_{CCCH}$ it is H-Cγ-Cβ-Cα.

The hybridisation of the α- and β-carbons is particularly significant,[118] such as the comparison of the experimentally measured $^{2}J_{CH}$ values for ^{13}C-enriched acetylene (49.4 ± 0.2 Hz), ethene (-2.4 ± 0.2 Hz) and ethane (-4.5 ± 0.3 Hz) reported by Lynden-Bell and Sheppard.[119] ^{1}H that couple through the sp^2 carbon of an aldehyde or enamine also have a characteristically large and positive values (20-30 Hz).[8]

Table I.3 A selection of typical values for $^2J_{CCH}$ depending on the coupling pathway.[22]

Coupling pathway	$^2J_{CCH}$ /Hz
H-C-C	-5-5
H-C=C	≤10
H-C≡C	40-60
H-C(=O)-C	20-25

For 3D structural determination (the investigation of configuration and/or conformation), the relationship of $^2J_{C(X)CH}$ to the dihedral angle (φ) between the H-C(β) and C(α)-X is of interest. DFT calculations of the FC contribution to $^2J_{CH}$ for acetaldehyde, ethylamine and ethanol can be approximated by Equation I.17, previously discussed with respect to $^2J_{HCH}$ with an electronegative group on the intermediate carbon (acetaldehyde A = -5.06, B = -2.15, C = 1.32, D = -0.18; ethylamine A = -3.10, B = -2.92, C = 0.49, D = -0.05; ethanol A = -3.36, B = -1.82, C = 0.12, D = -0.08). However, for $^2J_{CH}$ the dihedral angle (φ) is between H-C(β) and C(α)-X (Figure I.13) rather than the H-C and X-H bonds for $^2J_{HCH}$ (Figure I.11).[112]

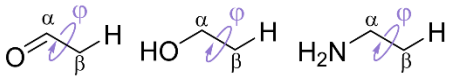


Figure I.13 Definition of dihedral angle (φ) for $^2J_{C(X)CH}$ in acetaldehyde, ethylamine and ethanol.

Recent additions to empirical $^2J_{CH}$ equations include further modifications to Equation I.17, which produced a generalised treatment of single α and β- substituents (H to Br) for sp^3 carbon-only coupling pathways by introducing substituent electronegativity (Δχ) and H-C-C bond angle corrections using DFT-calculations of $^2J_{CH}$.[120]

I.3.2.3. Three-bond proton-carbon scalar coupling: $^3J_{CH}$

The inclusion of an additional atom to the $^3J_{CH}$ coupling pathway further increases the possible combinations of hybridisations and number of substituent sites compared to $^2J_{CH}$. Following experimental evidence suggesting that a Karplus-like relationship existed between Φ (defined in Figure I.14) and $^3J_{CH}$,[121,122] Wasylishen and Schaefer[6] calculated the FC contribution to $^3J_{CH}$ with changing Φ for an HCC^{13}C fragment using the semi-empirical INDO (Intermediate Neglect of Differential Overlap) method. A Karplus-like relationship to Φ was confirmed and Equation I.13 fitted with the calculated data (A = 4.26, B = -1.00 and C = 3.56), which gave the first empirical equation for $^3J_{CCCH}$ prediction.

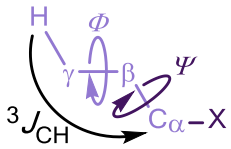


Figure I.14 $^{3}J_{CH}$ coupling pathway, defining the dihedral angles of interest (Φ and Ψ).

Aydin and Günther later sought to generate a more accurate empirical equation for purely hydrocarbon systems[123] using experimentally measured $^{3}J_{CD}$ ($^{n}J_{CH} = \gamma_H/\gamma_D \times {}^{n}J_{CD}$) for deuterated molecules with conformationally rigid structures (substituted norbornanes, fenchanes and adamantanes) and dihedral angles (Φ) calculated by molecular mechanics. The relationship between the experimental $^{3}J_{CH}$ values and calculated dihedral angles followed a Karplus-like equation (Equation I.13, A = 4.19, B = -0.27, C = 3.52), in agreement with that calculated by Wasylishen and Schaefer. The effects of methylation in the α-, β- and γ-positions were also investigated. No significant α-effects were found and although for Φ < 90° there were no significant β- or γ-effects, when Φ > 90° methyl substitution at the β-position decreased $^{3}J_{CH}$ by 0.9 ± 0.3 Hz and at the γ-position by 1.7 ± 0.5 Hz. To include the methyl substituent effects the coefficients in Equation I.13 were revised to A = 4.50, B = -0.87, C = 4.03.

The effect of substitution in the α-position on $^{3}J_{C(X)CCH}$ was studied in further detail by van Beuzekom, de Leeuw and Altona through a computational study of substituted propanes (X = H, CH_3, NH_2, OH or F).[7] As shown in Figure I.15, $^{3}J_{CH}$ was found to depend on both the dihedral angle between the coupled nuclei (Φ) and the CCCX dihedral angle (Ψ) as defined in Figure I.14.

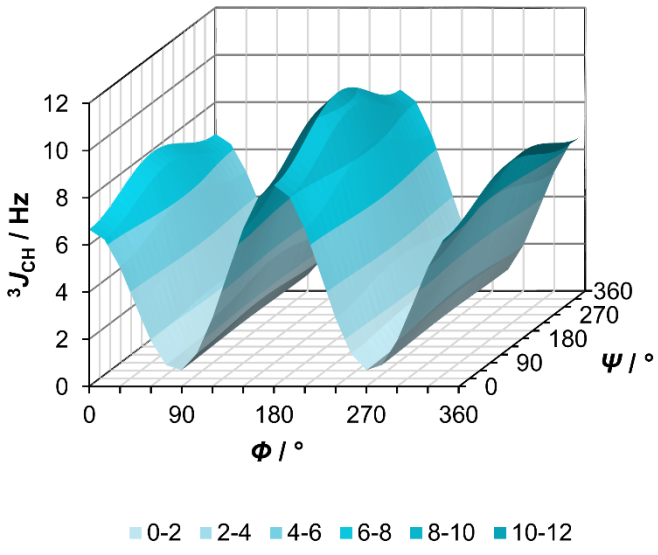


Figure I.15 $^{3}J_{CH}$ as a function of dihedral angles Φ and Ψ (Equation I.18) for an average X α-substituent (C_{00} = 4.55, C_{10} = -1.10, C_{20} = 3.83, C_{01} = -0.41, C_{11} = 0.16 and C_{21} = -1.38).

$$^{3}J_{\mathrm{CH}} = C_{00} + C_{10}\cos\Phi + C_{20}\cos 2\Phi + \cos\Psi\,(C_{01} + C_{11}\cos\Phi + C_{21}\cos 2\Phi)$$ Equation I.18

$^{3}J_{CH}$ as a function of two dihedral angles (Φ) and (Ψ) for α-substituent (X = H, CH_3, NH_2, OH or F).

It was found that $^{3}J_{CH}$ also depended on the electronegativity of the α-substituent and so the equation was further modified, Equation I.19.

$$^{3}J_{\mathrm{CH}} = (C_{00,0} + C_{01,0}\Delta\chi) + (C_{10,0} + C_{11,0}\Delta\chi)\cos\Phi + (C_{20,0} + C_{21,0}\Delta\chi)\cos 2\Phi + (C_{01,1} + C_{11,1}\cos\Phi + C_{21,1}\cos 2\Phi)\Delta\chi\cos\Psi$$ Equation I.19

Equation describing $^{3}J_{CH}$ in the presence of an α-substituent (X) as a function of two dihedral angles (Φ) and (Ψ) with a correction for the electronegativity of the substituent relative to hydrogen ($\Delta\chi$).[7]

The effect of substituent electronegativity on $^{3}J_{CH}$ was studied by Palermo, Riccio and Bifulco for β and γ-substituents for $^{3}J_{CCCH}$. Equation I.14 which accounts for intermediate substituent effects on $^{3}J_{HH}$ was modified to Equation I.20 and fitted using data from computational models of substituted butane and pentane molecules.

$$^{3}J_{\mathrm{CH}} = P_1\cos^2\Phi + P_2\cos\Phi + P_3 + \sum \Delta\chi_{H\pm}\left(P_4 + P_5\cos^2(\xi_i\Phi + P_6|\Delta\chi_{H\pm}|)\right) + \sum \Delta\chi_{C\pm}\left(P'_4 + P'_5\cos^2(\xi_i\Phi + P'_6|\Delta\chi_{C\pm}|)\right)$$ Equation I.20

Empirical equation from Palermo *et al.*[10] accounting for β and γ-substituent effects for $^{3}J_{CH}$. $P_{1\text{-}6}$ and $P'_{4\text{-}6}$ are constants. $\Delta\chi$ is the electronegativity of the substituent relative to hydrogen where H indicates γ-substituents and C β-substituents. ξ_i is 1 or -1 depending on the orientation of the substituent, as defined in Figure I.16.

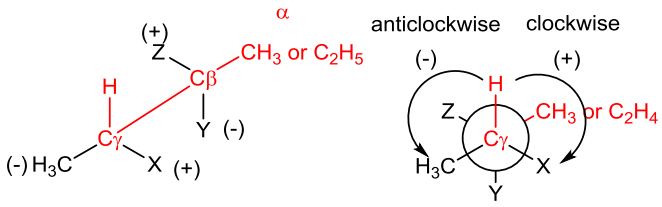


Figure I.16 Definition of positive and negative substituents in Equation I.20.

The focus of this overview of empirical $^{3}J_{CH}$ prediction has been generalised equations which have the broadest applicability. However, another approach is to target highly specialised equations for a single molecular fragment such as those included in reviews of structure determination of nucleic acids,[103] peptides/proteins (e.g. amide $^{3}J_{CNCH}$)[105,106] and carbohydrates

(e.g. glycosidic $^{3}J_{COCH}$)[107], these equations are therefore of use where the fragment repeats in a related biomolecule.

I.3.3. Scalar couplings between X and Y: $^{n}J_{XY}$

This section has focussed on the relationship between scalar coupling and 3D structure for $^{n}J_{CH/HH}$ as commonly applied to elucidating the structure of typical organic molecules that are rich in hydrogen and carbon. However, this discussion is not exhaustive, in 2000 Contreras and Peralta published the most comprehensive review to date on the angular dependence of $^{n}J_{XY}$ scalar couplings, which discussed bond/dihedral angle and substituent effects on $^{n}J_{XY}$ where X and Y included H, C, N, F, P, Sn.[112] A recently published program (MSpin-JCoupling)[124] provides a tool that gathers together several empirical equations used for small organic molecules, peptides and glycosides ($^{3}J_{HH}$, $^{3}J_{CH}$, $^{3}J_{HF}$, $^{3}J_{CC}$, $^{3}J_{NH}$ and $^{2}J_{CH}$) and predicts (population weighted) scalar couplings from inputted 3D molecular structures.

I.4. Summary and outlook

This thesis is primarily concerned with improving NMR-based methods for determining three-dimensional (3D) molecular structure in solution with a focus on $^{n}J_{CH}$.

The empirical $^{3}J_{CH}$ equations discussed in section I.3 were either highly specialised (targeting a single molecular fragment) or generalised (accounting for effects such as substituent electronegativity). However, the applicability of these general equations is hindered by the nature of the coupling pathway used; they are based on fragments containing sp^3 carbons. Therefore, they have limited use in molecules with coupling pathways containing sp^2/sp carbons or heteroatoms with any hybridisation state, as are frequently encountered in both natural products and pharmaceutics. The possible substitution patterns that are accounted for are similarly limited, and accounted only for the orientation and electronegativity of β- and γ-substituents[10] or single α-substituents[7]. It is therefore of practical importance to explore the impact of interactions between all α-, β- and γ-substituents and the dihedral angles relating them on $^{3}J_{CH}$ (starting in Chapter III using DFT-calculated $^{3}J_{CH}$).

The interpretation of the magnitude of any DFT-calculated variation in $^{3}J_{CH}$ (due to changes in the molecular structure) first requires an assessment of the accuracy of experimental measurements of $^{3}J_{CH}$. Therefore, Chapter II begins with an exploration of the accuracy of advanced 2D methods to measure $^{n>1}J_{CH}$ experimentally by comparison to one-dimensional (1D) coupled ^{13}C spectra for model compounds. It also includes a comparison of these experimentally measured scalar couplings to values calculated by DFT.

Chapter II. Accurate experimental measurement of $^{n}J_{CH}$

Parts of the work described in this chapter have been published in the peer-reviewed journal "Analyst" under the title "Accurate measurement of long range proton–carbon scalar coupling constants", authored by Claire L. Dickson, Charles D. Blundell, Craig P. Butts, Alice Felton, Alex Jeffreys and Zoltan Takacs.[125]

II.1. Introduction

Proton-carbon scalar coupling constants, $^{n}J_{CH}$, are a powerful tool in the elucidation of 3D structure, as discussed in Chapter I, in particular for parts of molecules with few ^{1}H where $^{n}J_{HH}$ or ^{1}H-^{1}H NOESY data may consequently be sparse. However, the interpretation of experimentally measured $^{n}J_{CH}$ in relation to 3D structure in a quantitative (e.g. comparison to DFT computed, or empirically predicted values) rather than qualitative or semi-quantitative manner (e.g. the *J*-based configurational approach[126,127]) requires a greater consideration of the accuracy of the experimental measurements.

Numerous methods have been developed to measure $^{n}J_{CH}$, as detailed in several recent reviews.[128-130] However relatively little consideration has been given to the accuracy of the values determined. Until recently only selected values across different techniques,[129,131] comparison between variations of related methods[132] or comparison to literature that had not been validated[133,134] had been investigated. In 2016 Pierens *et al.*[135] published a comparison of 59 $^{2,3}J_{CH}$ measured by 1D coupled ^{13}C and IPAP-HSQMBC spectra for a subset of four cinnamic acid derivatives. A set of twelve derivatives (Figure II.1) were then used to validate DFT[136] calculated $^{2,3}J_{CH}$ by comparison to those measured by IPAP-HSQMBC. The comparison of the two experimental methods showed an average deviation of 0.1 Hz between the 36 $^{3}J_{CH}$ and 0.23 Hz between the 23 $^{2}J_{CH}$. The couplings ranged between 3.0-10.0 Hz and 1.0-4.5 Hz for the $^{3}J_{CH}$ and $^{2}J_{CH}$ respectively.

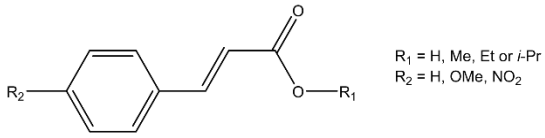


Figure II.1 Cinnamic acid derivatives studied by Pierens *et al.*[115]

In this thesis, it was of interest to extend this type of analysis to more structurally diverse molecules in addition to a wider range of NMR methods to establish accuracy, ease of analysis and experimental efficiency in comparison to coupled ^{13}C measured $^{n}J_{CH}$. The molecules strychnine and camphor were chosen as test molecules (Figure II.2). Both are rigid in nature which simplifies the comparison of experimentally measured and calculated $^{n}J_{CH}$, and have reasonable ^{1}H T1/T2 relaxation times (>0.4s for strychnine and >2.0s for camphor in $CDCl_3$), which do not limit the 2D NMR methods that can be tested. Strychnine also benefits from being

a commonly studied molecule by NMR, and therefore comparison to $^{n}J_{CH}$ reported in the literature was possible.[129,133,134,137,138]

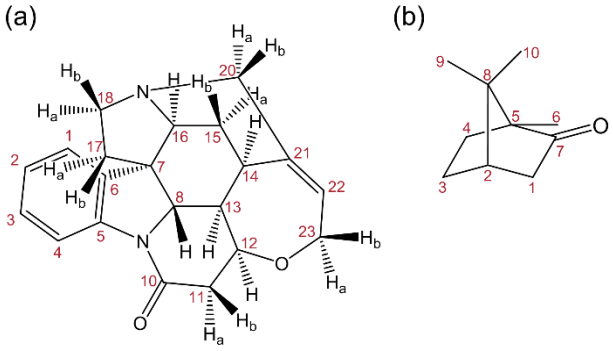


Figure II.2 Labelling system used for strychnine (a) and camphor (b). The germinal protons in camphor are distinguished using Ha (*pro-R*) and Hb (*pro-S*), Appendix 1 for full ^{1}H and ^{13}C assignment.

The intended use of $^{n}J_{CH}$ data includes the elucidation of the 3D structure of challenging ^{1}H-poor molecules and therefore techniques that cannot measure scalar couplings involving quaternary carbon centres were not considered, such as variants of the HETLOC,[139] HECADE,[140] and HSQC-TOCSY[141,142] experiments. Additionally, coupled ^{13}C spectra are not generally suitable for accurately determining the sign of $^{n}J_{CH}$ (except in some rare cases involving significant second order effects). Therefore, techniques that can measure couplings to quaternary carbons but included sign (HSQMBC-TOCSY,[143] selHSQMBC-COSY,[144] etc.) were not explored.

II.2. 1D coupled ^{13}C measurement of $^{n}J_{CH}$

The 1D gated ^{1}H decoupled ^{13}C experiment was used to establish an accurate “Gold Standard” set of $^{n}J_{CH}$ for comparison to other techniques; this method applied decoupling only during the relaxation delay and meant that a coupled ^{13}C spectrum could be recorded while benefiting from NOE enhancement. The complex multiplets present in coupled ^{13}C spectra required simulation and fitting procedures to extract $^{n}J_{CH}$. This was initially achieved for camphor using the spin simulation tool in MestReNova 9.0.1[145]. However, the MestReNova software does not have fitting functionality associated with the spin simulation tool and therefore the suitability of fit was assessed by eye (overlaying the simulated and experimentally measured spectra and comparing the peaks). The $^{n}J_{CH}$ were assigned by matching the size and sign of the experimental coupling constants to those calculated by DFT.

One challenge is the assignment of $^{n}J_{CH}$ to ^{1}H-^{13}C pairs from coupled ^{13}C spectra alone, in comparison to $^{n}J_{CH}$ the assignment of $^{n}J_{HH}$ in 1D ^{1}H spectra is relatively simple due to two signals sharing the same coupling. The low natural abundance of ^{13}C (1.07%[146]) means that ^{1}H spectra are dominated by ^{1}H nuclei attached to ^{12}C therefore and so $^{n}J_{CH}$ are generally not visible, except for $^{1}J_{CH}$ satellites that are free from overlap with other ^{1}H nuclei (Figure II.3).

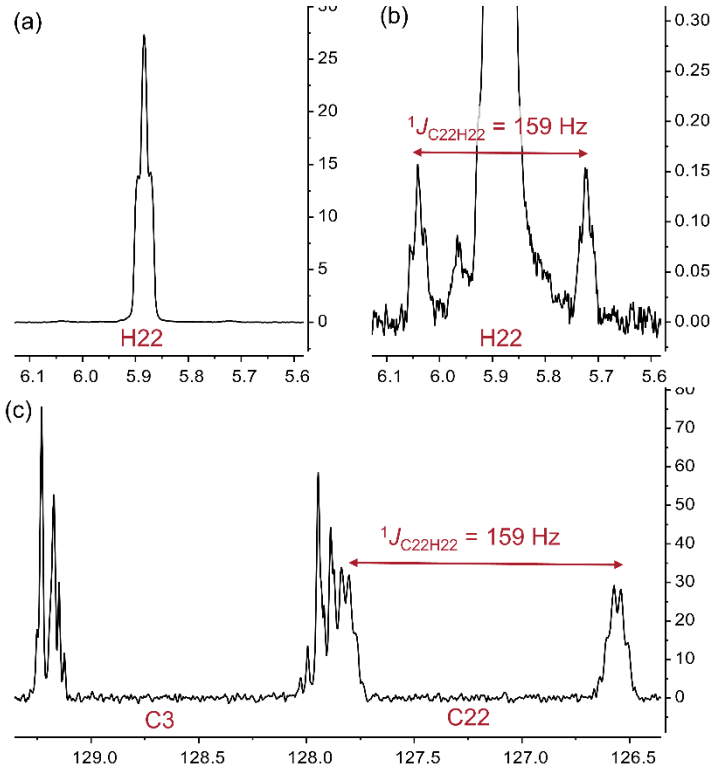


Figure II.3 Measurement of $^1J_{CH}$ using H22 (a/b) and C22 (c) in strychnine as an example.

The $^nJ_{HH}$ and ^{1}H chemical shifts obtained from simulated ^{1}H spectrum assisted in the assignment of $^nJ_{CH}$ in peaks dominated by second order effects such as those from ABX systems[43]. However, further confirmation of the assignment of $^nJ_{CH}$ in coupled ^{13}C spectra measured from 1D spectra was obtained for camphor by selectively decoupling ^{1}H resonances; in these spectra $^nJ_{CH}$ that are caused by the decoupled ^{1}H are absent. In cases where a high degree of coupling from the nine methyl ^{1}H (H6, H9 and H10) caused broad and complex peaks (Figure II.4c) in camphor, selective decoupling made possible the measurement of additional $^nJ_{CH}$ (Figure II.4a/b).

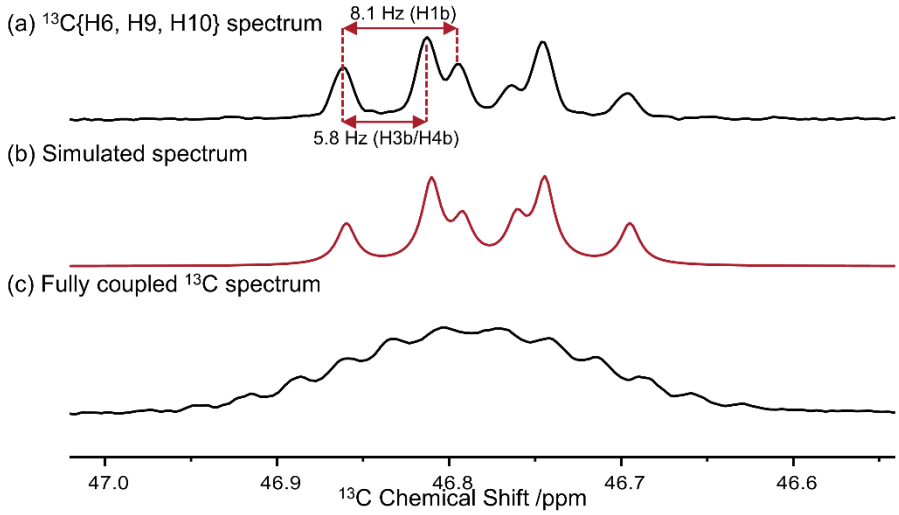


Figure II.4 Coupled ^{13}C spectra of C8 in camphor, (a) selective decoupling of the methyl ^{1}H (H6, H9 and H10), (b) simulated spectrum of (a) and (c) the fully coupled ^{13}C spectrum.

To explore the effect of different approaches in the analysis of the same coupled ^{13}C spectra four different analyses were performed for camphor (summarised in Table II.1, full datasets in Appendix 2, Table A.5). The first (a) used the MestReNova software and included simulations of selectively decoupled peaks and resulted in 22 measured coupling constants. The second (b) performed by Alexander Jeffreys[147], did not include the selectively decoupled data and instead included estimates of $^{1}J_{CH}$ for peaks where the $^{n>1}J_{CH}$ was too complex for a full simulation; this produced 21 $^{n}J_{CH}$. The third analysis (c), performed by Dr Zoltan Takacs (C4X Discovery)[148] used a GAMMA[149] based simulation method including fitting procedures, full details in Chapter VIII (section VIII.1.5). This method was capable of extracting coupling constants from more complex multiplet shapes in the fully coupled ^{13}C spectrum and measured 32 $^{n}J_{CH}$. Subsequently, this method was applied to the selectively decoupled ^{13}C spectrum (d) which did not include coupling from the methyl ^{1}H. In the absence of the methyl peaks, 30 $^{n}J_{CH}$ were measured. For all analysis methods (a-d) where the assignment of the ^{1}H corresponding to the coupling was ambiguous the best fit on comparison to DFT-calculated values was assumed to be the correct assignment.

Table II.1 Spin simulation of camphor coupled ^{13}C spectra. (a) MestReNova spin simulation including $\delta(^1H)$ and $^nJ_{HH}$ and simulations of selectively decoupled peaks. (b) MestReNova spin simulation including $\delta(^1H)$ and $^nJ_{HH}$ performed by Alexander Jeffreys[147]. (c) Dr Zoltan Takacs, C4X Discovery simulation[148]. (d) Zoltan Takacs, C4X Discovery, spin simulation of methyl ^1H decoupled spectrum[148].

Simulation	(a)	(b)	(c)	(d)
Software	MestReNova 9.0.1	MestReNova 9.0.1	C4X Discovery	C4X Discovery
Including selectively ^1H decoupled spectra?	Yes	No	No*	Methyl decoupled
Total $^nJ_{CH}$	22	21	32	30
Average error in $^nJ_{CH}$ parameters	N/A	N/A	0.7 Hz	0.7 Hz

*Selectively ^1H decoupled spectra informed the starting points of the fitting procedure.

There was very good agreement between these four methods with a mean standard deviation of 0.2 Hz between the values and a total of 45 measured couplings (including $^1J_{CH}$). However, it was clear that the fitting methods (c) and (d) of the simulated spectra (both fully coupled and selectively decoupled) allowed more confident measurement of a larger number of $^nJ_{CH}$ and therefore, this method was used in subsequent extraction of coupling constants for 1D spectra and, when necessary, slices of 2D spectra. The same simulation/fitting procedure was applied to strychnine, which has a range of complexity in the ^{13}C spectrum, for example the carbonyl C10 (Figure II.5a) was very simple and $^nJ_{CH}$ could be directly measured, while the aromatic C2 (Figure II.5b) was significantly impacted by second order (strong coupling) effects.

Overall, 55 $^{n>1}J_{CH}$ values were extracted from coupled ^{13}C spectra for strychnine and camphor (Appendix 2). These were subsequently used as “Gold Standard” values to determine the accuracy of the other 2D NMR methods discussed in this chapter. The 55 “Gold standard” scalar coupling constants excluded $^1J_{CH}$, $^{n>1}J_{CH}$ with high (HESSE) errors in the fitting and couplings <1 Hz where the (HESSE) error in the fitting is generally larger than the size of the $^{n>1}J_{CH}$. These errors, calculated by the HESSE algorithm[150], represent a maximum range of error in $^nJ_{CH}$ reflecting the peak linewidth and complexity.

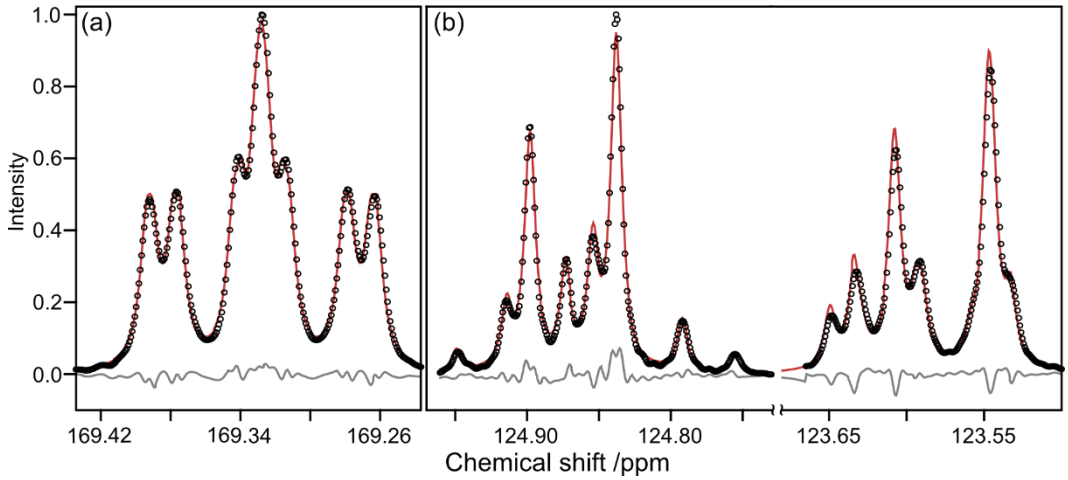


Figure II.5 Fitting of coupled ^{13}C spectrum for strychnine C10 (a) and C2 (b). The black circles are the measured spectrum, the red line the fitted spectrum and the grey line the absolute error. Reproduced from reference.[125]

II.3. Measurement of $^{n>1}J_{CH}$ in F1

Despite coupled ^{13}C spectra providing accurate $^{n}J_{CH}$, direct ^{13}C detection means these spectra suffer from low sensitivity, while the requirement for complex simulation/fitting procedures and assignment issues makes the technique less attractive for routine use than 2D methods that are generally ^{1}H detected and identification of the coupled nuclei is inherent in the F1/F2 chemical shifts. In these experiments, the simplest extraction of the $^{n>1}J_{CH}$ can be achieved from encoding $^{n>1}J_{CH}$ in the F1 dimension (Figure II.8). In this work, the ^{1}H-selective technique, EXSIDE[151] (a/b) and the broadband *J*-HMBC[152,153] (c/d) were examined as examples of this approach.

EXSIDE and *J*-HMBC both use *J*-scaling in the F1 dimension to avoid requiring a large number of t1 increments for achieving a high enough resolution in F1 to measure small couplings (for typical ^{13}C spectral widths >10,000 increments are required for <1Hz resolution). However, each of these methods achieves this scaling in a very different manner (full pulse sequences are shown in Figure II.7 and Figure II.6).

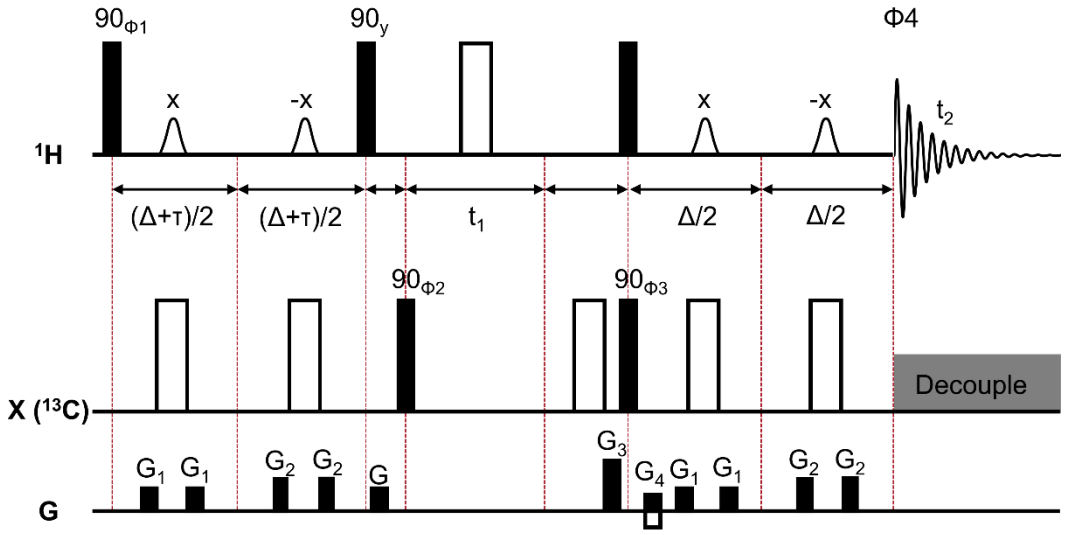


Figure II.6 The pulse sequence for EXSIDE. The filled bars represent 90_x and open bars 180_x pulses unless otherwise indicated while the curved pulses are band-selective. The delays are $\Delta = 1/(2J_{LR})$ and $\tau = N \times t_1$ where N is the chosen scaling factor. Phase cycle: Φ1= x,x,x,x,-x,-x,-x,-x; Φ2 = x,x,-x,-x; Φ3 = x,-x; Φ4 = x,-x,-x,x,-x,x,x,-x. Diagram based on reference[151].

The EXSIDE pulse sequence starts with a ^{1}H-selective INEPT using a DPFGSE (Double Pulsed Field Gradient Spin Echo) that includes two delay periods of (Δ+τ)/2. The length of τ determines the size of the scaling factor, N, and is a function of t1, τ = N×t1. A balance of two factors determines the size of N; the larger N, the fewer t1 increments required to obtain a given resolution in the measured coupling constant, which therefore shortens experiment times. However, larger N increases the INEPT delay period for magnetisation transfer and hence nuclei experience more relaxation, which can substantially reduce signal intensity. The length of the INEPT evolution delay Δ is chosen by considering the typical frequency of long range $^nJ_{CH}$ such that $\Delta=1/(2\times J_{LR})$, thereby leading to maximum efficiency in the polarisation transfer. However, the ^{1}H-selective nature of the technique limits the application to single ^{1}H or groups of mutually uncoupled ^{1}H, requiring the recording of multiple spectra (2-4 hours of experiment time per spectrum) to investigate a full molecule.

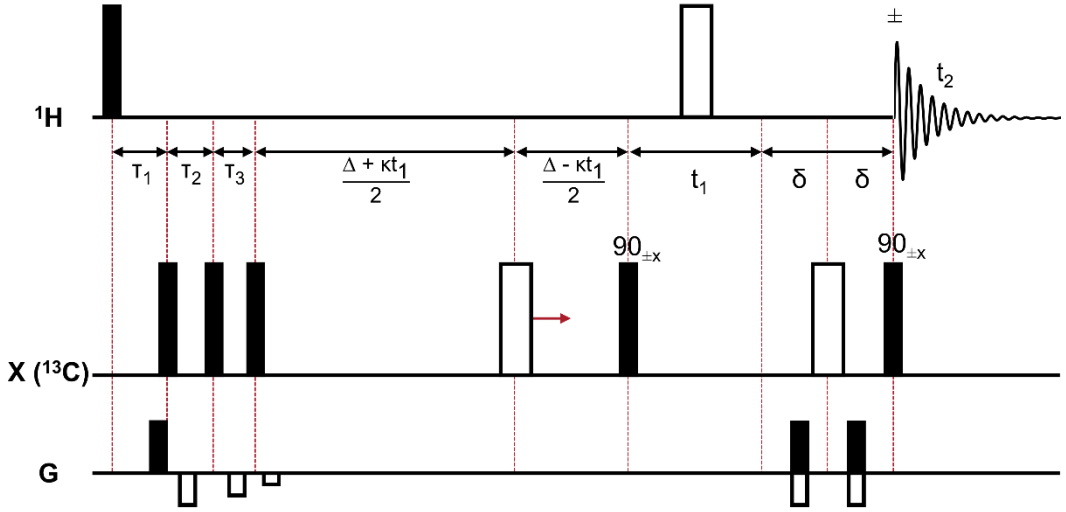


Figure II.7 The *J*-HMBC pulse sequence. The delays $\tau_{1\text{-}3}$ in the low-pass *J*-filter are set using a typical maximum (J_{max}) and minimum (J_{min}) $^1J_{CH}$ as follows: $\tau_1 = 1 / [2 \times (J_{min} + 0.07 \times (J_{max} - J_{min}))]$; $\tau_2 = 1 / [J_{max} - J_{min}]$; and $\tau_3 = 1 / [2 \times (J_{max} + 0.07 \times (J_{max} - J_{min}))]$. δ is the gradient delay and $\kappa = \Delta / t_1^{max}$. The relative strengths of the gradients in the low pass filter are 7:-4:-2:-1 and the gradient echo are 5:-3 (echo), -3:5 (anti-echo). Diagram based on reference[152].

The *J*-HMBC, unlike EXSIDE, is a broadband technique which achieves *J*-scaling in F1 by incrementing the position of a 180° ^{13}C pulse by κ×t1 within a fixed period of length Δ prior to t1. The scaling factor κ is therefore determined by the F1 spectral width, number of t1 increments and length of $\Delta(=1/(2\times J_{LR}))$. In contrast to EXSIDE, the choice of J_{LR} is set by the size of smallest coupling to be measured rather being set by the size of a typical $^nJ_{CH}$ value. As shown in Figure II.8(b/d) the lineshape of the correlations in *J*-HMBC were more complex than those arising from the EXSIDE. It was also noted that the lineshape was dependent on the parameters chosen, so changing the scaling factor κ by manipulating the F1 spectral width and number of t1 increments while keeping Δ constant varied the lineshape quite substantially.

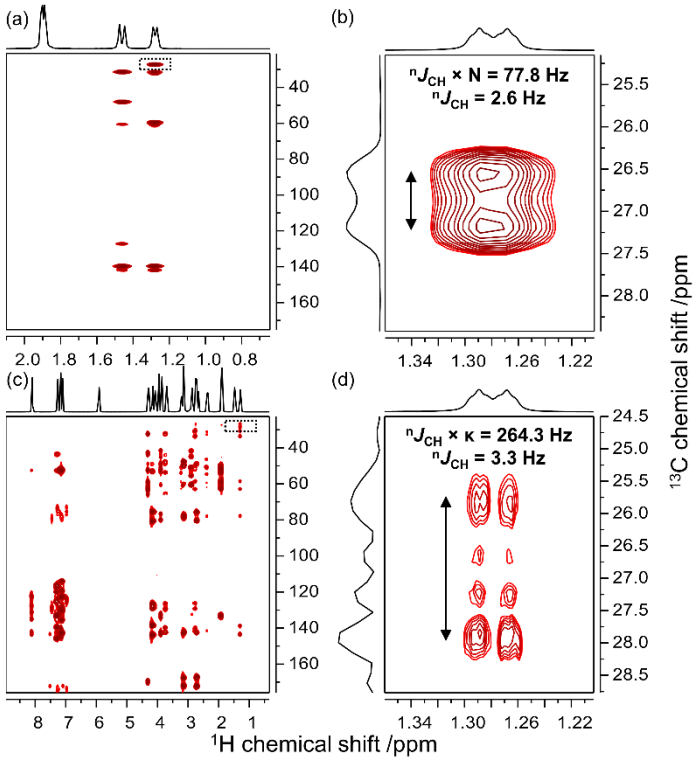


Figure II.8 EXSIDE (a/b) and *J*-HMBC (c/d) spectra of strychnine. (a) Full spectrum with J_{LR} = 6 Hz and N = 30, (b) enlarged section. (c) Full spectrum with J_{LR} = 1 Hz and κ = 79, (d) enlarged section. Amended from reference [125].

II.3.1. EXSIDE

Comparison of $^nJ_{CH}$ measured by coupled ^{13}C to those measured by other 2D methods was performed by calculating two key statistics: the mean absolute deviation (MAD) and standard deviation (SD) as defined in the experimental section (Chapter VIII). The $^nJ_{CH}$ measured by EXSIDE with a scaling factor, N = 30, and J_{LR} of 6 Hz showed a much higher MAD/SD than expected, 0.48/0.55 Hz across 27 values measured for strychnine and camphor. This high error was systematic, small couplings (<6 Hz) measured by EXSIDE appeared as smaller than expected, whilst large couplings (>6 Hz) were larger than expected, Figure II.9.i.

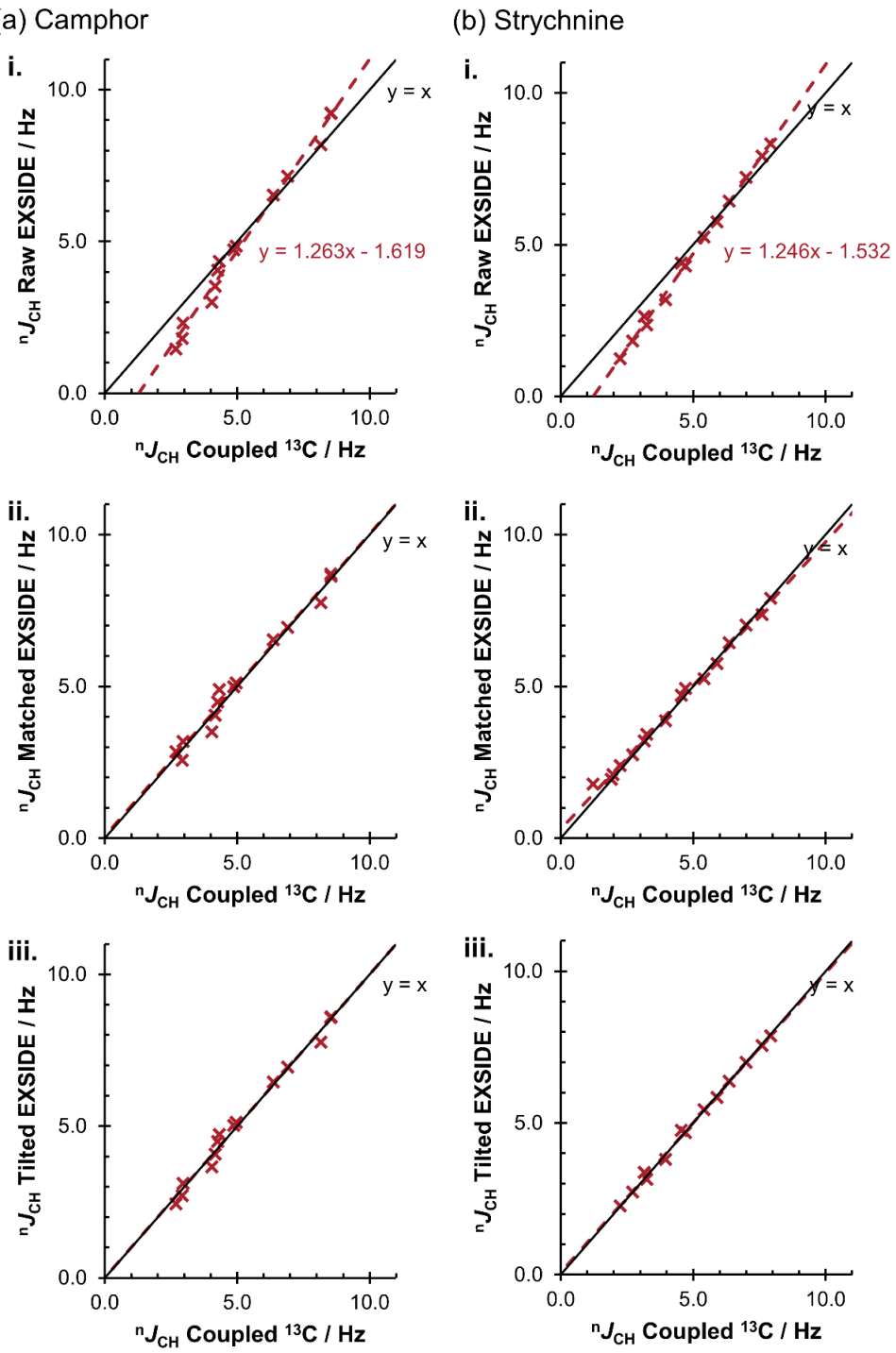


Figure II.9 Comparison of EXSIDE to coupled ^{13}C for (a) camphor and (b) strychnine. Where i. raw EXSIDE data with J_{LR} = 6 Hz and N = 30, ii. matched across J_{LR} = 2-10 Hz, N=30 (as described in the text), iii. tilted EXSIDE with J_{LR} = 6 Hz and N = 30 using gradient and intercept for (a)i. (as described in the text).

However varying J_{LR} (and thus the INEPT delay) from 4-10 Hz showed that the size of the splitting measured (Figure II.10) was dependant on J_{LR} and accurate $^{n}J_{CH}$ could therefore only be achieved when J_{LR} was close to $^{n}J_{CH}$. This technique is referred to as 'matching', where

EXSIDE spectra were recorded for J_{LR} = 2-10 Hz in 2 Hz steps, all $^{n}J_{CH}$ were extracted from the spectra, and those within 1 Hz of J_{LR} used to record that spectrum were included in the matched dataset. If the measured $^{n}J_{CH}$ was between two J_{LR} and more than 1 Hz from either, then the two $^{n}J_{CH}$ values were averaged across the two experimental values. This resulted in an MAD/SD that was significantly improved, 0.19/0.24 Hz across 30 $^{n}J_{CH}$, for strychnine and camphor (Figure II.9.ii).

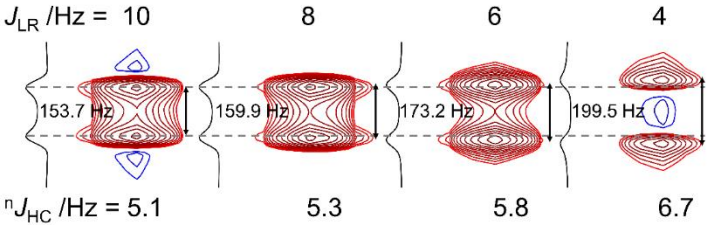


Figure II.10 EXSIDE spectra of strychnine C20-H22 with variable J_{LR}, where the scaling factor N = 30. Reproduced from reference [125].

Despite the accuracy achieved by matching, recording an additional 4 spectra for each ^{1}H (or group of mutually uncoupled ^{1}H) in a molecule requires a considerable amount of extra experiment time: 8-16 hours of experiment time per ^{1}H, compared to 2-4 hours for a single J_{LR}. Therefore, a simple 'tilting' procedure was implemented to correct for the systematic error arising from deviation from J_{LR}. Equation II.1 was used to correct the $^{n}J_{CH}$ measured for strychnine (J_{LR} = 6 Hz, N= 30) by using the gradient and intercept determined from plotting $^{n}J_{CH}$ measured for camphor (J_{LR} = 6 Hz, N= 30) against the coupled ^{13}C measured $^{n}J_{CH}$ (Figure II.9(a)i.).

$$^{n}J_{CH}(\text{tilted}) = \frac{^{n}J_{CH}(J_{LR}) - \text{intercept}(J_{LR})}{\text{gradient}(J_{LR})} \qquad \text{Equation II.1}$$

Where gradient(6 Hz) = 1.263 and intercept(6 Hz) = -1.619 Hz for camphor.

This significantly improves the accuracy of the strychnine $^{n}J_{CH}$ (Figure II.9.iii.), reducing the MAD/SD to 0.08/0.11 Hz across 13 values. Repeating this procedure for all J_{LR} measured using the camphor data alone gave the gradients and intercepts detailed in Table II.2 which also includes gradients and intercepts for the strychnine and camphor data combined (Figure II.11).

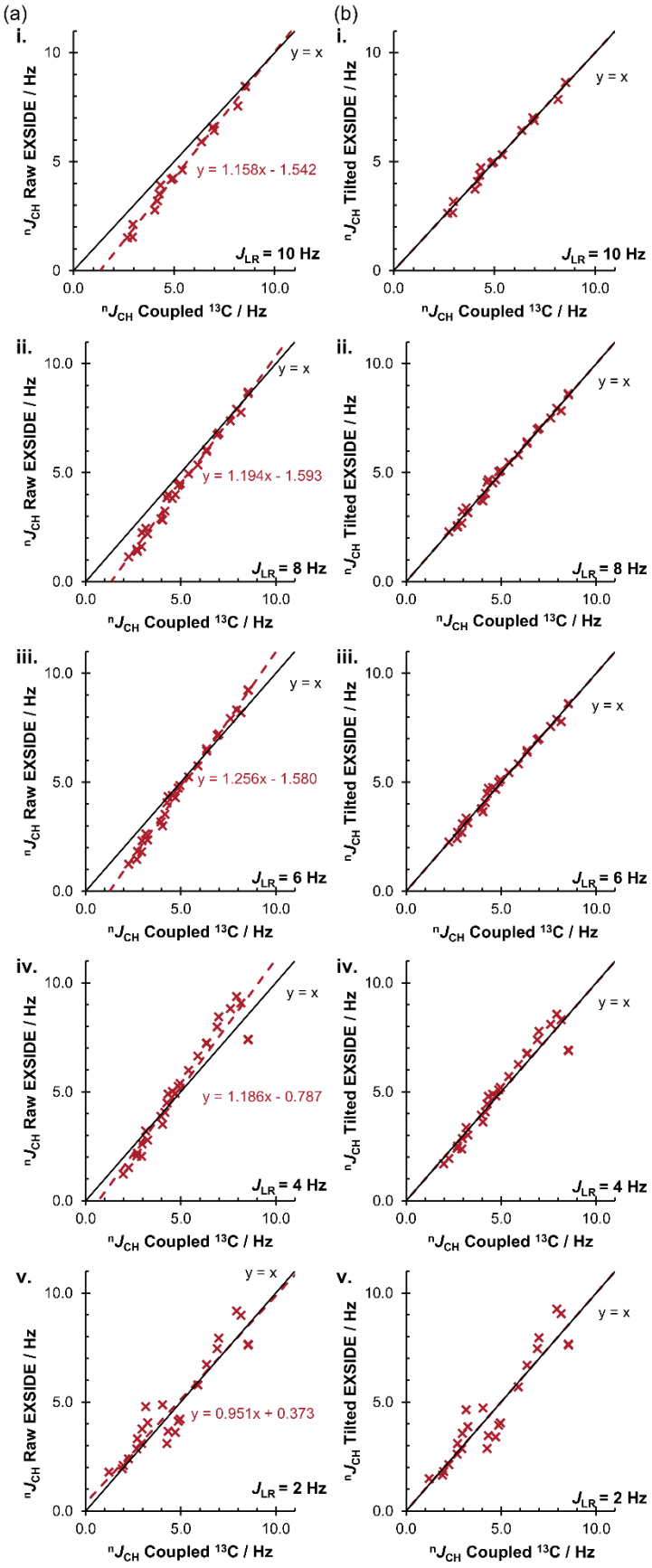


Figure II.11 Determining parameters for Equation II.1 with strychnine and camphor data combined. (a) raw EXSIDE data and (b) tilted EXSIDE with Equation II.1 for i.-v. with J_{LR} = 2-10 Hz and N = 30.

Table II.2 Gradients and intercepts from camphor data only (a) and combined strychnine and camphor data (b) for Equation II.1.

	J_{LR} /Hz	10	8	6	4	2
(a) camphor	gradient(J_{LR})	1.162	1.187	1.263	1.039	0.936
	intercept(J_{LR})	-1.548	-1.534	-1.619	-0.225	0.231
(b) camphor and strychnine	gradient(J_{LR})	1.158	1.194	1.256	1.186	0.951
	intercept(J_{LR})	-1.542	-1.593	-1.580	-0.787	0.373

Finally, to confirm the general nature of these corrections the camphor correction determined with N = 30 and J_{LR} = 8 Hz was applied to data for strychnine that had been recorded with the same J_{LR} but with a smaller scaling factor, N = 15 (Figure II.12). This achieved a similar accuracy to the N = 30 dataset, MAD/SD were 0.09/0.10 across 11 $^{n}J_{CH}$.

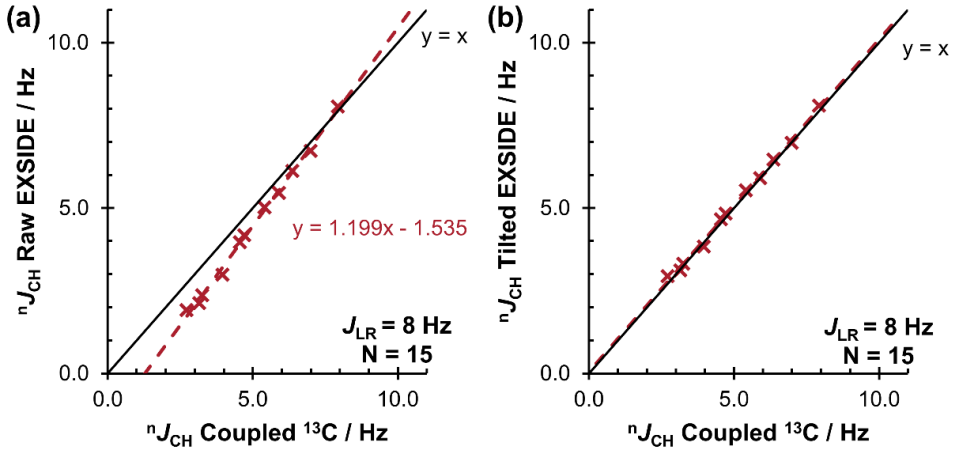


Figure II.12 Comparison of $^{n}J_{CH}$ measured by EXSIDE and coupled ^{13}C for strychnine where (a) is raw and (b) is tilted with Equation II.1 (camphor only, J_{LR} = 8 Hz).

II.3.2. *J*-HMBC

The ^{1}H-selective nature of EXSIDE limits the $^{n}J_{CH}$ which can be measured when coupled ^{1}H are close in chemical shift (approximately <40 Hz separation), however the *J*-HMBC method provides a broadband alternative where, in principle, couplings can be measured for all ^{1}H in the molecule in one spectrum. A typical EXSIDE or *J*-HMBC spectrum requires 1-4 hours (depending on the exact parameterisation). Therefore, the broadband *J*-HMBC results in an experimental time saving of a factor of the number of ^{1}H selections required by EXSIDE. However, this comes at a price – namely from increased distortion of the F1 lineshape, Figure II.8(d). The 'correct' coupling constant was assumed to be encoded in the splitting between the largest two peaks in the F1 projection of the corresponding correlation, however substantial secondary bands within the correlation were often as large as the 'correct' peaks. Except for

H2-C7 in camphor (Figure II.14, discussed below) it was found that the outer bands corresponded to an appropriate $^{n}J_{CH}$ value.

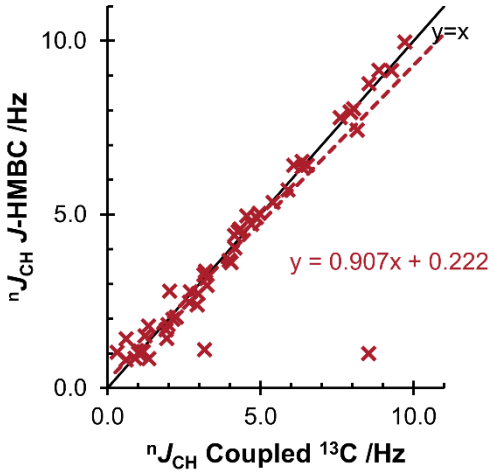


Figure II.13 Comparison of $^{n}J_{CH}$ measured by *J*-HMBC and coupled ^{13}C for strychnine (κ = 79, J_{LR} = 1 Hz) and camphor (κ = 56, J_{LR} = 1 Hz).

The large majority of $^{n}J_{CH}$ values measured by *J*-HMBC for strychnine and camphor (Figure II.13) were found to be reasonably accurate compared to coupled ^{13}C (MAD/SD, 0.28/0.44 Hz) without the systematic errors found for EXSIDE; however, this accuracy came with a significant caveat. The substantial lineshape distortion, as shown in Figure II.8(d), complicates the process of identifying the 'correct' peaks from which to measure the requisite coupling constant. When studying a completely new molecule, this lack of clarity in how to extract $^{n}J_{CH}$ from a given correlation makes it difficult to be confident in any one $^{n}J_{CH}$.

An extreme example of this is the substantially erroneous $^{n}J_{CH}$ measured for H2-C7 in camphor, it was measured as 1.0 Hz by *J*-HMBC in comparison to 8.4-8.7 Hz by DFT and all other experimental methods reported herein (Table A.6). The corresponding correlation can be seen in Figure II.14 where the EXSIDE (a) shows the expected splitting while *J*-HMBC (b) suggests a misleadingly small value. When this value is included the MAD/SD was 0.42/1.14 Hz, the largest of all the methods tested, due to the difficulty in interpreting the peak shape the value was excluded from the reported MAD/SD of 0.28/0.44 Hz.

This erroneous value was also observed in the corresponding constant-time variant of the *J*-HMBC recorded with the same experimental parameters[152]. Changing the *J*-scaling factor *κ* to 30 was observed to mostly restore the lineshape to one which more closely reflected the expected $^{n}J_{CH}$ value (7.5 Hz). This dependence of the extractable coupling constant on the parameterisation of the experiment raised significant concerns over the reliability and confidence which can be placed on $^{n}J_{CH}$ values extracted from *J*-HMBC even though the majority were accurate.

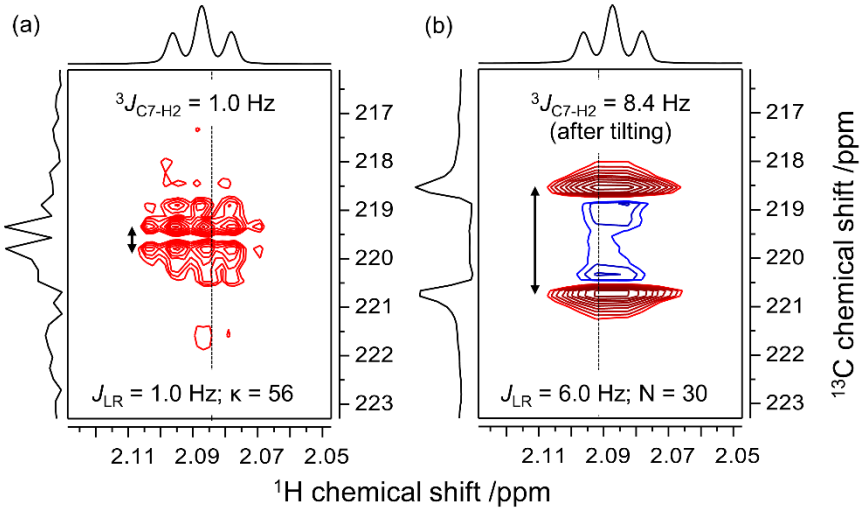


Figure II.14 *J*-HMBC spectrum (a) of camphor C7-H2 measured $^nJ_{CH}$ with high error and the corresponding peak in EXSIDE (b). The ^{1}H spectrum is shown along f2 and the dashed line indicates the position of the trace along f1. Reproduced from reference [125].

II.4. Measurement of $^{n>1}J_{CH}$ in F2

Almost all recently reported (non-TOCSY) methods for measuring $^nJ_{CH}$ in the direct (F2) dimension of 2D NMR spectra are variants of three fundamental approaches, namely HMBC, refocussed and non-refocussed HSQMBC. There are two broad approaches to extracting the $^nJ_{CH}$ from the spectra: either lineshape analysis of F2 multiplets in the HMBC/HSQMBC spectra or IPAP analysis utilising sum/difference spectra obtained from two separate in-phase (IP) and antiphase (AP) spectra. These two approaches to extracting $^nJ_{CH}$ are based on similar underlying NMR sequences, but comparison of the relative merits of each technique for measuring accurate $^nJ_{CH}$ was of interest.

II.4.1. Lineshape fitting

The co-evolution of the ^{1}H-^{1}H and ^{1}H-^{13}C scalar couplings during t_2 of HMBC and HSQMBC spectra creates complex F2 multiplets from which the $^nJ_{CH}$ must be extracted; this complexity makes accurate measurement of $^nJ_{CH}$ directly from the multiplet difficult, and consequently lineshape fitting is a requirement for non-trivial molecules. Methods based on F2-evolution of $^nJ_{CH}$ have been reported in a number of cases based on HMBC,[137,154] non-refocussed HSQMBC,[155,156] and pure in-phase HSQMBC (PIP-HSQMBC)[157]. Due to the shorter pulse sequences the non-refocussed methods are more sensitive than refocussed for protons with short relaxation times but the lineshapes are substantially more complicated and the extraction of coupling constants is usually more challenging. Recently a F2-homonuclear decoupled PSYCHE CPMG-HSQMBC[158] was reported to overcome the complexity of multiplets in these two cases but the severe loss in sensitivity limits the applicability to concentrated samples.

In a refocused HSQMBC, the $^{n}J_{CH}$ coupling appears as an in-phase addition to the proton multiplet. The key benefit to this technique is that the in-phase couplings can be manipulated to give cleaner lineshapes that are more easily fitted. Although accurate extraction of the couplings from these multiplets still requires simulation and lineshape fitting in almost all cases (and thus require prior accurate measurement of all the contributing $^{n}J_{HH}$) as well as a well-resolved multiplet. In order to make this simulation and extraction of $^{n}J_{CH}$ as easy as possible, ^{1}H-^{1}H *J*-modulation should be suppressed as it can perturb the peak shape. The PIP-HSQMBC[157] pulse sequence achieves this with a Keeler-type Zero Quantum filter at the end of the refocusing INEPT period.

The substantial challenge of measuring $^{n}J_{CH}$ values using refocused HSQMBC (without ZQ filters) or simple HMBC methods[138,154] is illustrated in Figure II.15 using the IP F2 slices of H8-C5 in strychnine. The refocused HSQMBC (b) and HMBC (c) showed significant lineshape variations that were dependant on J_{LR} which cannot readily be fitted while the PIP-HSQMBC (a) gives reliably in-phase lineshapes for H8-C5 as a function of the evolution period ($\Delta = 1/(2 \times J_{LR})$).

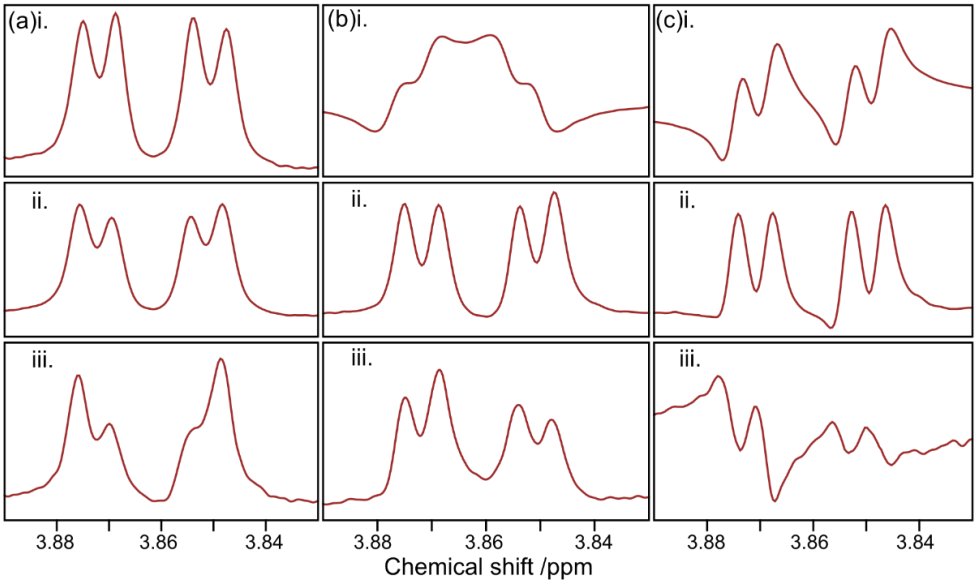


Figure II.15 H8-C5 of strychnine ($^{n}J_{CH}$ = 3.2 Hz) IP slices of PIP-HSQMBC (a), refocused HSQMBC (b) and HMBC (c) at different coherence transfer times, i, ii, iii, J_{LR} = 4, 6, 8 Hz respectively. The PIP-HSQMBC spectra used a 20ms 60kHz CHIRP pulse and 3% in the ZQ filter. Reproduced from reference [125].

Suppression of the $^{1}J_{CH}$ residual signals was achieved with TANGO excitation[159] and GBIRD[160] elements in both the initial and reverse INEPT periods. The introduction of extra elements such as the ZQ-filter, TANGO and/or BIRD further reduced the intensity of the faster relaxing peaks of these long refocused HSQMBC sequences, however they were necessary for successful lineshape fitting of the complex multiplets generated by the refocused HSQMBC. This

sequence (the refocussed HSQMBC including ZQ filter, TANGO and BIRD elements) is referred to as 'PIP-HSQMBC' and the sequence without these elements as 'refocused HSQMBC' throughout. The HMBC and refocussed HSQMBC methods gave data that was challenging to fit reliably; therefore, only the fitting of PIP-HSQMBC will be considered further.

Lineshape fitting of the in-phase PIP-HSQMBC spectrum was performed by Dr Zoltan Takacs and used the same GAMMA[149] based simulation method described for the analysis of coupled ^{13}C spectra. This fitting procedure extracted only 20 $^{n}J_{CH}$ values (out of 101 $^{n}J_{CH}$ >1 Hz by DFT) for strychnine. The lineshape fitting was necessary in order to extract almost any coupling constants and accurate fitting of the in-phase lineshapes was often not possible, *i.e.* the fitting was unstable, or the fitting errors were substantially greater than the desired 0.5 Hz accuracy. For example, H22 of strychnine showed a very broad apparent triplet ^{1}H resonance, and the H22-C23 peak in the PIP-HSQMBC shows an ~7 Hz coupling, but simulation and fitting of this peak gives a high error range (7.3 ± 1.1 Hz) due to uncertainties in the underlying linewidth and multiple small $^{n}J_{HH}$ that contribute to the broadened lineshape. In this instance, H22 was relatively isolated in the ^{1}H spectrum and thus can be selectively excited and homodecoupled by HOBS-decoupling[132,161] with a substantial improvement in the error range for the fitting as well as improved sensitivity (6.8 ± 0.3 Hz). However, in addition to requiring selective excitation, which is not always possible, the HOBS decoupling artificially broadens the underlying linewidth (to ~2.5 Hz in this case), especially where narrow selection bandwidths (and thus long selective pulses) are used, which may make very small couplings hard to measure accurately. Selective excitation also means that generally couplings to only one proton at a time can be measured, so the use of HOBS substantially increases the required experiment time if all the coupling constants for a molecule are to be measured, hence it is only recommended for resolving challenging multiplets.

When they could be fitted, the $^{n}J_{CH}$ values extracted from the in-phase PIP-HSQMBC gave accurate results (0.18 Hz MAD, 0.14 Hz SD) in comparison to the values established from coupled ^{13}C spectra. While this suggests that reliable values extracted by this method are highly accurate, it should be noted that these MAD/SD values were calculated from only six couplings that were pre-filtered by fitting quality and that could be measured by both IP PIP-HSMQBC and coupled ^{13}C, Figure II.16.

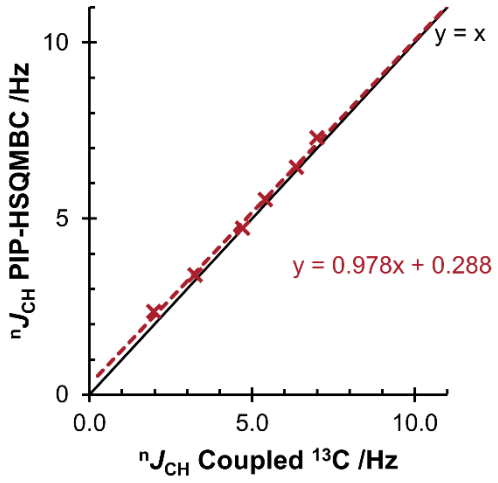


Figure II.16 Comparison of strychnine and camphor $^nJ_{CH}$ measured by lineshape fitting of PIP-HSQMBC, J_{LR} = 6 Hz.

II.4.2. IPAP analysis

However, another way that HSQMBC and HMBC methods can be used to extract the $^nJ_{CH}$ is to incorporate the spin state selective (in-phase antiphase, IPAP) principle, described by Parella *et al*.[154] In this type of experiment two separate datasets are recorded: one with the long-range scalar coupling in-phase (IP) and one with it antiphase (AP). The IP and AP experiments performed here were interleaved but can also be recorded separately. The sum and the difference of the two FIDs is formed in the time domain prior to Fourier transformation of each. An overlay of the sum and difference spectra then allows $^nJ_{CH}$ to be extracted from the offset between multiplets in F2 in each spectrum, rather than from the splittings within the multiplets. This is shown in the F2 traces in Figure II.17 for the large C10-H11b scalar coupling in strychnine and the small C8-H1a scalar coupling in camphor.

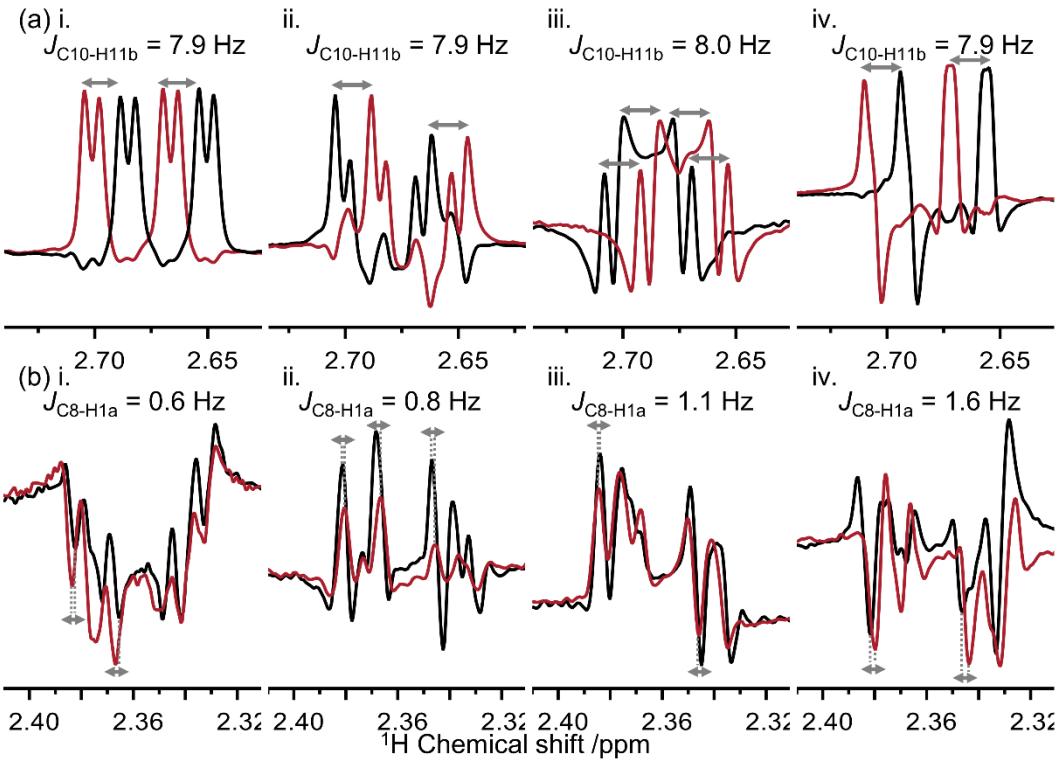


Figure II.17 IPAP analysis of C10-H11b of strychnine (a) and C8-H1a of camphor (b) applied to i. PIP-HSQMBC, ii. refocused HSQMBC, iii. accordion-HSQMBC, iv. HMBC. Where the sum is black and the difference in red and J_{LR} = 6 Hz for all except iii. Where J_{LR} = 3-8 Hz. Reproduced from reference [125].

In principle, this makes it substantially easier to extract couplings because it does not rely on having resolvable lineshapes within the multiplets nor on spectrum simulation and fitting procedures; however, it does assume that the multiplet shapes of the sum and difference spectra are comparable, which was not always found to be the case. The IPAP approach is much less sensitive to ^{1}H *J*-modulation than PIP-HSMQBC alone, since the modulation is the same in the two separate IP and AP experiments and thus has a consistent effect on both the sum/difference spectra and so does not affect the measured offset. There is therefore less benefit to incorporating a ZQ filter when using IPAP, unlike for lineshape analysis of the refocussed HSQMBC. If major imbalances occur between the two datasets, then the introduction of a post-acquisition scaling factor in the sum and difference step (IP±(k×AP)) can help to resolve this issue.[131] This does require that multiple sum/difference spectra are generated to extract the maximum number of coupling constants but does not require multiple experimental datasets to be acquired and so is relatively time-efficient.

This reduced sensitivity to lineshape distortion can be seen by comparison of the IPAP analysis of PIP-HSQMBC (Figure II.17i.) and refocused HSQMBC spectra (Figure II.17ii.), where it is clear that the PIP-HSMQBC data has improved lineshape. While the refocused HSQMBC is still interpretable in principle (and gives a similar $^nJ_{CH}$ if measured from the splitting between the main positive peaks) examination of the lineshapes reveals that the sum/difference

lineshapes are not identical. An even starker example is provided by an Accordion[162,163] variant of the HSQMBC developed by Dr Zoltan Takacs (Figure II.17iii.) whereby very substantial lineshape distortion occurs, but because it arises equally in both sum and difference spectra the corresponding coupling constant can still be extracted.

It is useful to note that the measurement of small $^{n}J_{CH}$ values (<2 Hz) is nearly always challenging due to insensitivity (so the multiplets are very weak) and substantial distortion of the multiplets from ^{1}H-^{1}H modulation. This latter point is clearly demonstrated in Figure II.17(b) where all IP and AP datasets show severe distortion, however offsets between these sub-spectra can still provide reasonable estimates of their magnitudes in each case. A number of methods have been proposed to improve sensitivity to these small $^{n}J_{CH}$ values, such as HSQMBC/HMBC-COSY,[154] which provide increased numbers of correlations in this $^{n}J_{CH}$ range, particularly for the very smallest (<1 Hz) couplings.

The IPAP analysis of HSQMBC and HMBC data substantially simplified the extraction of couplings, where the sum/difference lineshapes are similar, Figure II.17, then overlaying the spectra and measuring $^{n}J_{CH}$ takes just a few seconds and requires no substantial expert treatment. Even imperfect lineshapes can be analysed with more confidence in most cases, illustrated in Figure II.18(b) for HMBC. However, in some cases (Figure II.18(a), refocused HSQMBC) the lineshapes of the sum and difference spectra do not allow any certain extraction of the $^{n}J_{CH}$ value.

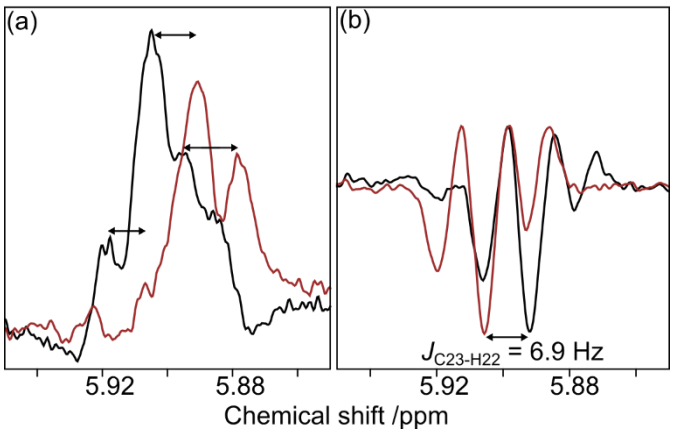


Figure II.18 IPAP analysis of H22-C23 in strychnine. (a) refocused HSQMBC (b) HMBC. Where J_{LR}= 8 Hz, sum (black) and difference (red) of IP and AP spectra. Reproduced from reference [125].

The effect of this lineshape dependence on accuracy can also be seen in the variation of measured $^{n}J_{CH}$ from the offset between different pairs of lines within the sample multiplet. This is illustrated in Figure II.19 for H1a-C3 of camphor and H23b-C21 of strychnine. In the latter case the multiplets can be interpreted as four peaks (which is the correct interpretation) or five (which is not correct, the smallest peak in each multiplet appears to be a sum/difference

artefact), so the user must be wary of measurements in distorted lineshapes such as this. It is also clear that the sum/difference lineshapes (red vs black) are not the same in each case, and so the precise value of $^{n}J_{CH}$ that is measured will depend on which pair of peaks the user selects to measure between, for example in the case of H23b-C21 (strychnine) this leads to a range of ±1.0 Hz in the extracted coupling constants depending on the lines selected. In such cases, where it was felt a reasonable estimate could be made, the $^{n}J_{CH}$ value reported is the average value of the various measured splittings.

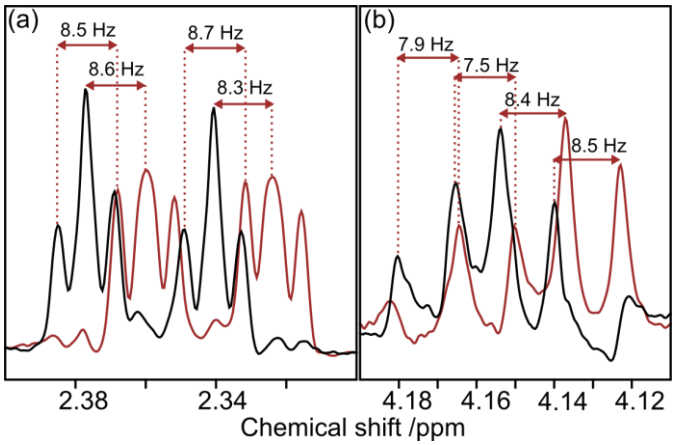


Figure II.19 Showing the spread of $^{n}J_{CH}$ in sum (black) and difference (red) for PIP-HSQMBC. (a) H1a-C3 of camphor, (J_{LR} = 6 Hz) (b) H23b-C21 of strychnine, (J_{LR} = 8 Hz). Reproduced from reference [125].

In some cases, these difficulties meant that correlations could not be confidently analysed. Even so, IPAP analysis provided high numbers of extracted $^{n}J_{CH}$ values for strychnine and camphor: 89, 74, and 74 out of 143 for IPAP HMBC, refocused HSQMBC and PIP-HSQMBC respectively, and 96 for IPAP analysis of the accordion HSQMBC, the highest of all methods except the *J*-HMBC.

The variability of lineshape between sum and difference spectra is the greatest practical drawback for IPAP-based method and prevents the reliable measurement of $^{n}J_{CH}$ values in some cases. In particular, the variability of the $^{n}J_{CH}$ values when measured from different peaks of the overlaid multiplets (Figure II.19) is a substantial source of any inaccuracy in this method; this can be minimised by calculating $^{n}J_{CH}$ as the average of all splittings measured between the sum/difference multiplets. However, when the sum and difference sub-spectra give comparable lineshapes, IPAP analysis gives reasonable agreement with coupled ^{13}C data (0.35 Hz MAD, 0.40 Hz SD for HMBC; 0.33 Hz MAD, 0.53 Hz SD for refocused HSQMBC; 0.22 Hz MAD, 0.34 Hz SD for PIP-HSQMBC), Figure II.20. These findings are in line with those reported recently by Pierens *et al.*[115] who found average deviations between coupled ^{13}C and IPAP analysis of refocused HSQMBC data of up to 0.2 Hz.

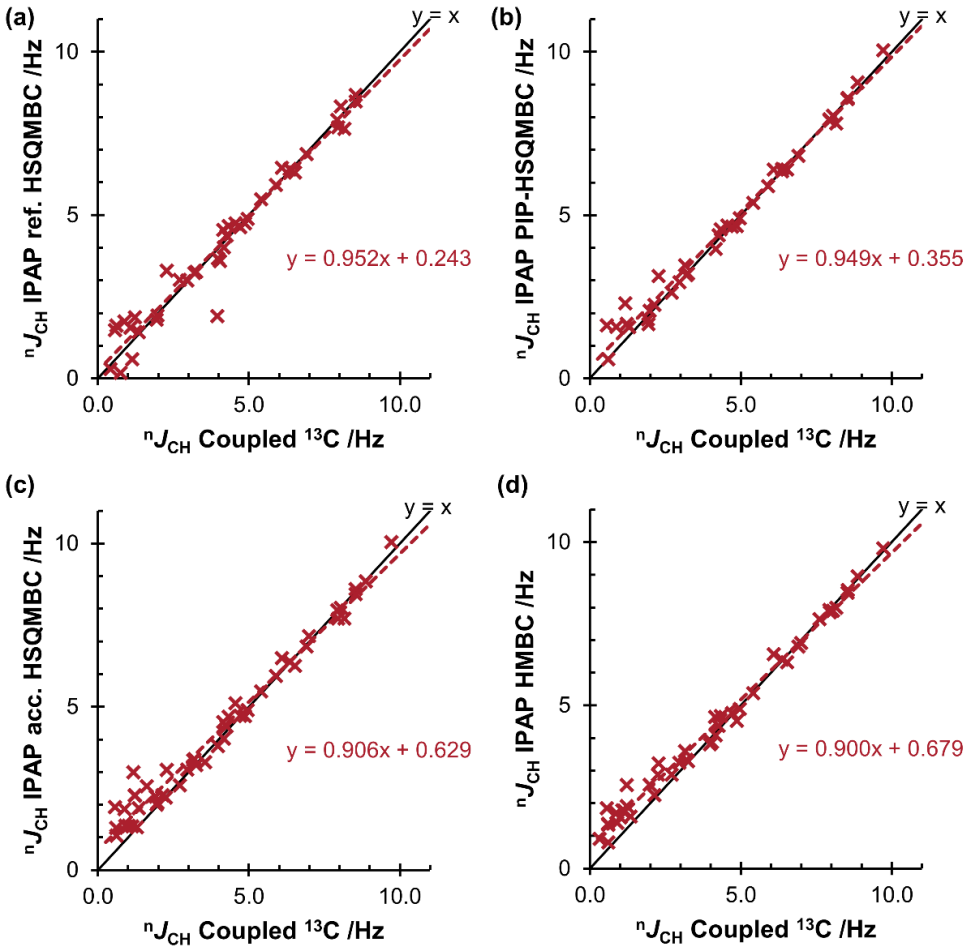


Figure II.20 Comparison of strychnine and camphor $^nJ_{CH}$ measured by IPAP analysis of HSQMBC/HMBC variants (a) refocused HSQMBC, (b) PIP-HSQMBC, (c) accordion-HSQMBC, (d) HMBC. Where J_{LR} = 6 Hz (a,b,d) or 3-8 Hz (c).

II.5. DFT-calculated and Literature $^nJ_{CH}$

It was also of interest to explore the accuracy of DFT-calculated $^nJ_{CH}$ and those reported in the literature. $^nJ_{CH}$ calculation was performed using Gaussian 09[78] starting with a stepwise geometry optimisation of strychnine and camphor, first with molecular mechanics (MM) with the Uniform Force field (UFF), then with density functional theory (DFT) with B3LYP/3-21G, then mPW1PW91/6-31G (d,p) and finally mPW1PW91/6-311G (d,p), (see Appendix 3 for Cartesian coordinates). NMR calculations were performed with mPW1PW91/6-311G (d,p) using the GIAO method and including total scalar coupling constants, consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms. DFT calculations were performed using the IEFPCM (integral equation formalism polarizable continuum model) solvent model for chloroform.

The comparison of DFT-calculated $^nJ_{CH}$ values to those obtained from coupled ^{13}C experiments demonstrated that the DFT method used systematically underestimated long range $^nJ_{CH}$ by

~6% and $^{1}J_{CH}$ by ~9% (Figure II.21). This was in line with previous findings for 88 $^{1}J_{CH}$ from small (<10 heavy atoms) organic molecules calculated with a combination of various functionals and basis sets (35 combinations) that attributed the ~5 Hz error to rovibration contributions not accounted for in the equilibrium $^{1}J_{CH}$ calculated.[164] A systematic percentage increase[165] in DFT-calculated $^{n}J_{CH}$ values was therefore applied, prior to comparison with experimental data; however, a relatively large mean absolute deviation (MAD, 0.38 Hz) and standard deviation (SD, 0.49 Hz) remained for the long range $^{n}J_{CH}$. This suggests that there is room for improvement in computational methods in order to match the quality of data available from experimental methods for moderately complex organic molecules; this is explored in more detail in Chapter VI.

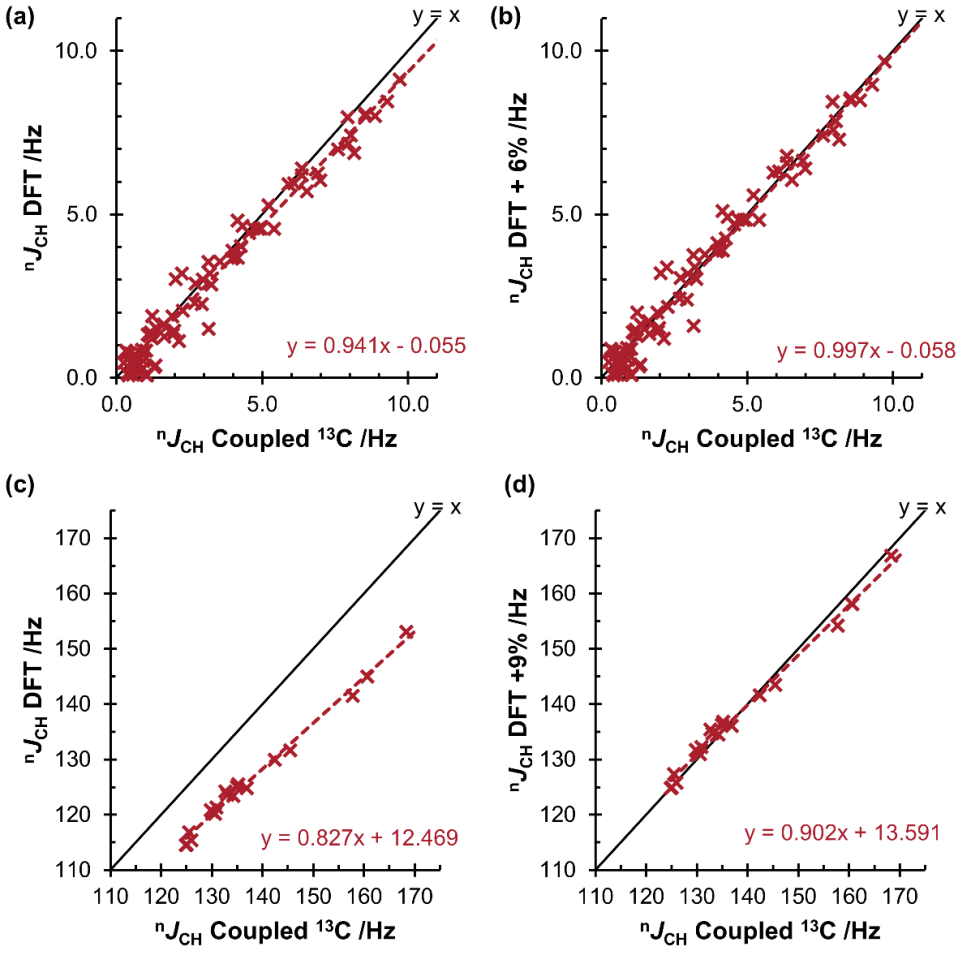


Figure II.21 Comparison of DFT-calculated $^{n}J_{CH}$ (mPW1PW91/6-311G(d,p), IEFPCM(chloroform)) and coupled ^{13}C $^{n}J_{CH}$, where (a/b) long range $^{n}J_{CH}$ and (c/d) $^{1}J_{CH}$.

It was also of interest to highlight the accuracy of $^{n}J_{CH}$ data from literature sources. Table A.4 contains $^{n}J_{CH}$ coupling constants for strychnine averaged across those found in historical reports using a range of techniques (sel-EXSIDE, HMBC fitting, HETLOC, HSQC-TOCSY, HSQC HECADE, GHSQMBC, HSQMBC, G-BIRD-HSQMBC, *J*-res-HMBC-2, *J*-IMPEACH-MBC, psHMBC, 3D HSQC-TOCSY, IPAP HSQC-TOCSY, IPAP-HSQMBC)[129,133,134,137,138,166]

and demonstrates that the reliability of these values is relatively low when they were compared to the coupled ^{13}C data measured here. The MAD of the literature data compared to coupled ^{13}C was 0.60 Hz (SD 0.91 Hz) although this was dominated by anomalous values measured from the 1D ^{1}H-detected spectra reported by Blechta *et al*.[166] This sequence generates antiphase doublets by using simultaneous WALTZ decoupling and a soft ^{13}C selective pulse followed by a BIRD module to achieve polarisation transfer to ^{1}H with an 8×8 Hadamard matrix in order to record couplings to multiple ^{13}C in one experiment with a sensitivity improvement of √8. Removing these values and consequently limiting the data to 2D techniques reduced this to MAD/SD of 0.48/0.67 Hz, this was still larger than the methods used in this thesis.

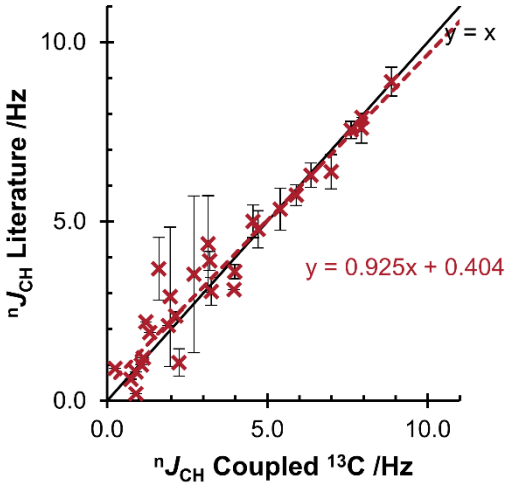


Figure II.22 Comparison of $^{n}J_{CH}$ for strychnine reported in the literature,[129,133,134,137,138] where the error bars indicate the SD between the different methods.

II.6. Summary

As summarised in Table II.3, the accurate measurement of 55 $^{n}J_{CH}$ couplings from coupled ^{13}C spectra allowed the assessment of the ease, efficiency and accuracy of recent F2- and F1-based 2D methods for measuring $^{n}J_{CH}$ in well-behaved model compounds, strychnine and camphor. For these compounds, where chemical exchange, line broadening/relaxation and spectrum overlap are not limiting, methods based on full spin-system simulation and/or F2 line-shape fitting (coupled ^{13}C, HSQMBC) are still found to be extremely time intensive for analysis and do not allow the accurate recovery of a large percentage of possible $^{n}J_{CH}$ values (<40%).

The IPAP-based analysis of HSQMBC and HMBC data were found to be much more robust and efficient, providing the largest number of the expected $^{n}J_{CH}$ values in a relatively short amount of experimental and analysis time. IPAP-based methods also allowed measurement of $^{n}J_{CH}$ with good accuracy (mean deviations of <0.5Hz from coupled ^{13}C data), although care must be taken when extracting $^{n}J_{CH}$ values from heavily distorted lineshapes.

Table II.3 Comparison of different methods to measure $^{n}J_{CH}$.

	DFT	Coupled ^{13}C[c]	IPAP accordion HSQMBC	IPAP PIP-HSQMBC	IPAP HMBC	IPAP refocussed HSQMBC	Tilted EXSIDE	*J*-HMBC	PIP-HSQMBC[f]	Lit.[h]
No. of $^{n}J_{CH}$	143[b]	55	96	74	89	74	73	98	20	86
% of 143 $^{n}J_{CH}$ measured		38	67	52	62	52	51	69	19	85
MAD (SD) /Hz[a]	0.38 (0.49)		0.33 (0.45)	0.22 (0.34)	0.35 (0.40)	0.33 (0.52)	0.08 (0.11)[d]	0.28 (0.44)[e]	0.18 (0.14)[g]	0.48 (0.67)
Exp. time per $^{n}J_{CH}$ /min		30	3	7	6	6	41	2	14	
Range of analysis time per $^{n}J_{CH}$ /min		Up to 72 hours	1-20	1-20	1-20	1-20	1-3	1-10	15-25	

[a] Mean absolute deviation/standard deviation (MAD/SD) calculated by comparison to coupled ^{13}C values.
[b] $^{n}J_{CH}$ >1 Hz after 6% correction.
[c] The reported values were taken from full matrix spin simulations and fitting of coupled ^{13}C spectra by preference. Where these lineshapes could not be simulated effectively due to their complexity, the values are reported from simulation and fitting of selectively decoupled ^{13}C spectra.
[d] MAD/SD calculated from 'tilted' EXSIDE values applied to strychnine data only, 13 values.
[e] MAD/SD value ignores the highly erroneous H2-C7 value. Including H2-C7 gives MAD/SD of 0.42/1.14 Hz.
[f] Data reported for lineshape analysis of PIP-HSQMBC for strychnine only.
[g] MAD/SD for PIP-HSQMBC analysis were calculated from only six values.
[h] Average $^{n}J_{CH}$ for strychnine found in literature.[129,133,134,137,138,166]

The ^{1}H-selective homonuclear decoupled *J*-scaled F1 evolution of $^{n}J_{CH}$ in experiments such as EXSIDE provides the simplest spectra for measurement of $^{n}J_{CH}$ but are extremely experimentally time-consuming when couplings for multiple ^{1}H are required. The accuracy of raw EXSIDE-based methods is relatively low due to mismatches between the fixed evolution delay and the range of $^{n}J_{CH}$ values, but this can be substantially improved by the simple data-tilting outlined by Equation II.1. This 'tilted' EXSIDE approach provided the most accurate $^{n}J_{CH}$ data from a 2D-experiment in this thesis. The much less experimentally time-consuming *J*-HMBC, which also employs F1 *J*-scaling, provides rapid access to $^{n}J_{CH}$ values. However the complex lineshapes make confident extraction of accurate values substantially more challenging and in one case an entirely erroneous $^{n}J_{CH}$ value was encoded in the F1 splitting of the corresponding correlation.

Consequently, where reliable and accurate extraction of $^{n}J_{CH}$ values is required, the IPAP-based methods would be recommended for the routine global analysis in larger studies and tilted EXSIDE analysis in non-expert environments or where particularly high accuracy of few couplings are required.

Chapter III. Empirical estimation of $^{3}J_{CH}$: the fragment-based approach

III.1. Introduction: the scope of the problem

In Chapter II the experimentally measured $^{n}J_{CH}$ data were used to assess the accuracy of $^{n}J_{CH}$ calculated using DFT in addition to comparing different experimental techniques (Section II.5). However as discussed in Section I.3, $^{n}J_{CH}$ can also be rapidly predicted empirically from the molecular structure. The measured $^{n}J_{CH}$ data in Chapter II can therefore also be used to assess the accuracy of empirical methods for predicting $^{n}J_{CH}$.

A wide variety of literature equations have been developed with a focus on prediction of three-bond ^{1}H-^{13}C scalar coupling constants, $^{3}J_{CH}$,[8,104-107,112,167] which are analogous to the Karplus equation for predicting $^{3}J_{HH}$[4,5]. The dihedral angle (Φ) between the coupled nuclei, as shown in Figure III.1, is the dominant geometric parameter in literature equations predicting $^{3}J_{CH}$. However, Karplus noted in 1963 for $^{3}J_{HH}$ that there are many other structural factors such as hybridisation, bond length, bond angle and substituent electronegativity that have significant effects on the magnitude of the coupling, concluding that: '*Certainly with our present knowledge, the person who attempts to estimate dihedral angles to an accuracy of one or two degrees does so at his own peril*'.[5]

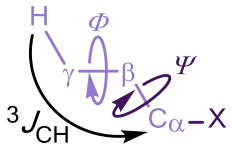


Figure III.1 $^{3}J_{CH}$ coupling pathway, defining the dihedral angles of interest (Φ and Ψ).

Numerous works have put substantial effort towards improving the accuracy of this approach for $^{3}J_{CH}$, for example with highly specialised equations targeting a single molecular fragment such as those included in reviews of structure determination of nucleic acids,[103] peptides/proteins[105,106] and carbohydrates[107]. An alternate approach is the use of generalised equations accounting for effects such as substituent electronegativity[7,10]. However, the applicability of these general equations is hindered by the nature of the coupling pathway used; they are based on fragments containing sp^3 carbons. Therefore, they have limited use in molecules with coupling pathways containing sp^2/sp carbons or heteroatoms with any hybridisation state, as commonly found in natural products and pharmaceutical molecules. The possible substitution patterns that are accounted for are similarly limited, for example the equation generated by Palermo *et al.*[10] accounted only for the orientation and electronegativity of β and γ-substituents (Equation I.20) on sp^3-hybridised carbons. However, earlier work from van Beuzekom *et al.* (Equation I.19)[7] noted that α-substituents could have significant effects

depending on the orientation to the coupling pathway, determined by the dihedral Ψ (Figure III.1). It is therefore of interest to explore the impact of interactions between all α, β and γ-substituents and the dihedral angles relating them on $^3J_{CH}$.

III.1.1. Generating empirical equations

Commonly, literature equations have been generated in one of two ways (summarised in Figure III.2). In the first,[123] a series of molecules are identified which contain $^3J_{CH}$ with the coupling pathway and substituent patterns of interest. The relevant $^3J_{CH}$ are then measured by NMR experiment and paired to the corresponding structural data (Φ, Ψ, …). Conformationally rigid molecules are frequently chosen such that a single set of geometric parameters are responsible for the measured $^3J_{CH}$, therefore the molecule can either be modelled by computation or an X-ray crystal structure measured to obtain the structural data.

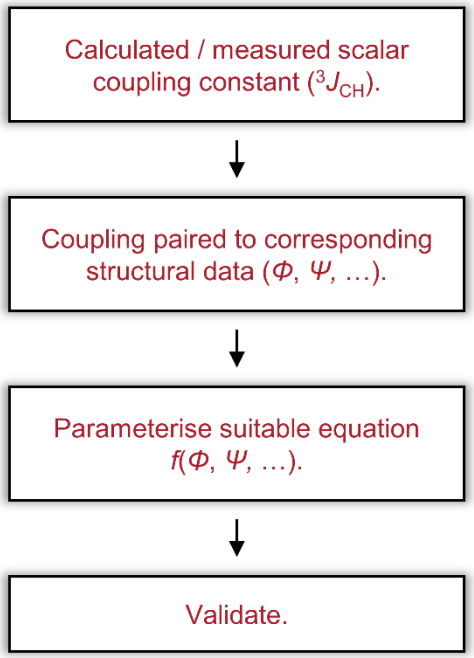


Figure III.2 Generalised process for generating empirical equations for predicting scalar coupling constants.

This set of $^3J_{CH}$ and the corresponding structural data can then be used to parametrise an equation with suitable behaviour, such as the Karplus equation (Equation I.13), the initial form of which approximates the results derived using valence-bond theory of the contact electron-spin contribution to the coupling of an HCCH fragment.[4] The performance of the equations is then confirmed with a set of $^3J_{CH}$ from molecules which were not used for parametrisation.

$$^{3}J_{\mathrm{HH}} = A + B\cos\Phi + C\cos 2\Phi$$ Equation I.13

Repeated for clarity. Where A, B and C are constants and A = 4.22, B = -0.50 and C = 4.50 Hz for sp^3 hybridised carbons with a C-C bond length of 1.543 Å.[5]

Another common method, employed by Palermo *et al.*[10] and van Beuzekon *et al.*[7], to generate their generalised equations uses both computationally (DFT and semi-empirical methods respectively) determined scalar coupling constants and structural data. In this approach a fragment that represents the target coupling pathway and substituent pattern is generated computationally and the geometric parameter of interest varied; the simplest example for $^{3}J_{CH}$ being a molecule of propane with the dihedral angle Φ varied in steps from 0 to 360°. An NMR calculation is performed for each geometric variation of the fragment to obtain $^{3}J_{CH}$. This process is performed with all combinations of coupling pathway or substituent pathway desired and the data used to parametrise an equation with suitable behaviour. Validation generally follows with experimental data from a set of molecules. In the cases of Palermo *et al*. and van Beuzekon *et al.* an additional step prior to validation was included, where the equations were generalised for additional structural variations between the fragments explored. This produced two general equations dependant on the substituent electronegativity of β and γ-substituents[10] or an α-substituent[7].

This second fragment-based approach has two main advantages: first, any coupling pathway or substituent pattern of interest can be generated computationally, avoiding the need to identify conformationally rigid molecules containing that fragment and secondly, the fragment can be generated with any set of geometric parameters desired, ensuring an even distribution of $^{3}J_{CH}$ data without biasing the equations generated to more common structural features. However, this approach is limited by the accuracy of the computational methods used for geometry optimisation and calculation of the scalar coupling constants. The number of calculations needed to describe the relationship between $^{3}J_{CH}$ and the geometric parameters of interest may also be large, potentially requiring a prohibitive level of computational resources.

In this thesis, a fragment-based approach similar to that of Palermo *et al.* and van Beuzekom *et al.* was chosen to generate $^{3}J_{CH}$ with fixed structural features for an array of substituents and coupling pathways. The sheer scale of sets of fragments needed for this approach indicates why the general empirical equations desired for different coupling pathways and substituent effects are not available. When limited to coupling pathways containing only carbon nuclei (Figure III.3), 35 different pathways result by combining the possible $sp/sp^2/sp^3$ hybridisation states with internal or external double/triple bonds along with distinguishing

aromatic from vinylic double bonds. These different pathways explode to over 130,000 possible fragments (Table III.1) when considering even a very limited subset of the typical substituents found in organic molecules: five substituents connected by a single bond (H, CH_3, OH, NH_2, F); three substituents connected by a double bond (CH_2, O, NH); or two substituents connected by a triple bond (CH, N). To scan the two dihedral angles Φ and Ψ in steps of 30° requires 338 calculations per fragment, although fewer calculations are needed where rotation about Φ and/or Ψ is not possible. If the CPU (central processing unit) time required is approximated to an average of 1 hour per calculation, the generation of empirical $^3J_{CH}$ equations for the subset of fragments described would take over 4,700 years of CPU time (Table III.1).

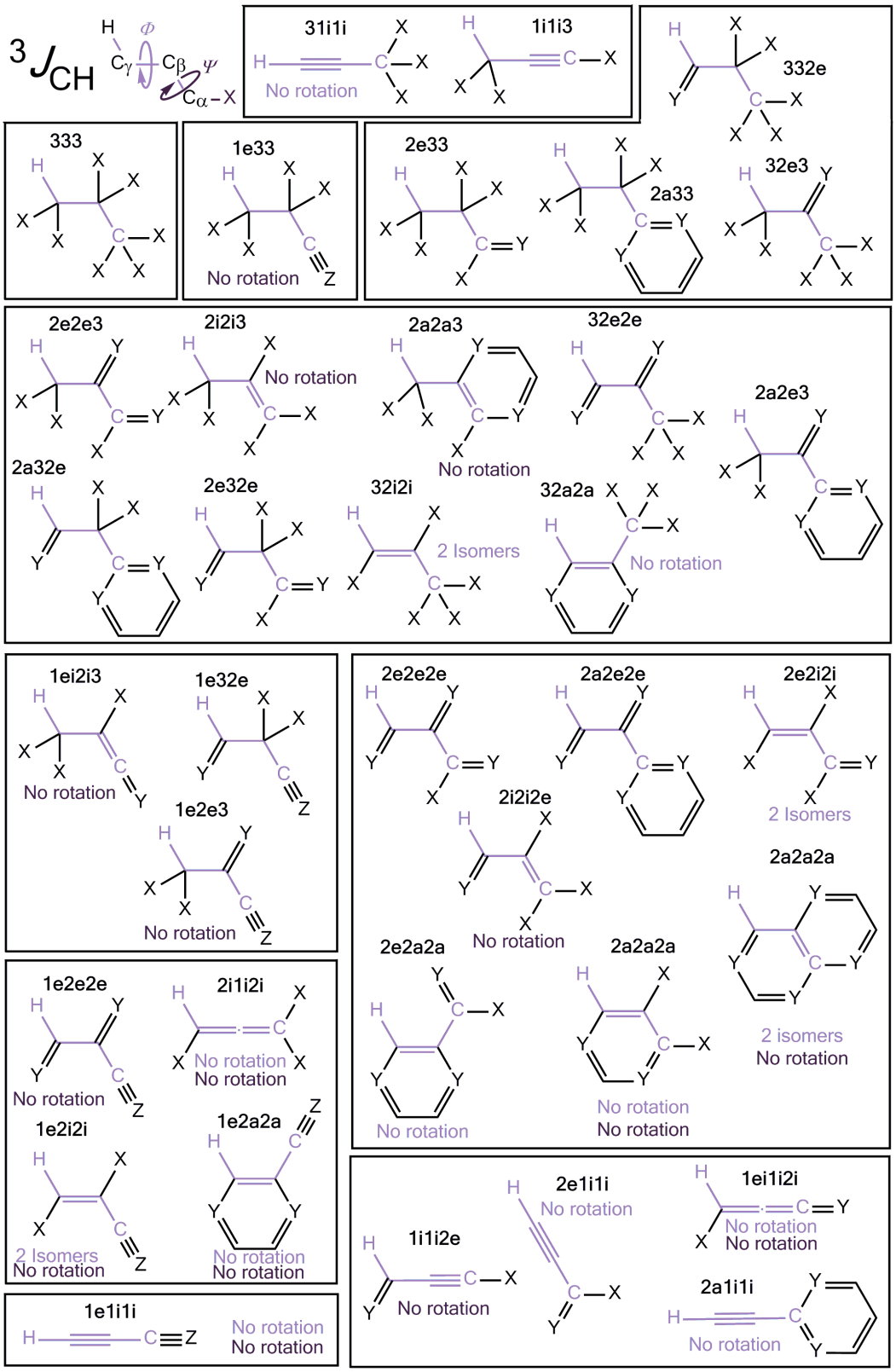


Figure III.3 Potential combinations of hybridisation giving rise to different carbon-only coupling pathways. Label format: αβγ, where sp^3 (3), sp^2 (2) and sp (1); internal multiple bond (i), external multiple bond (e) and aromatic (a). Substituents which form single bonds are indicated by X, double/aromatic bonds Y and triple Z. To indicate aromatic fragments a 6-membered ring is used, but in the case of heterocyclic compounds the ring size might necessarily vary.

Table III.1 Estimate of number of combinations of a limited subset of substituents for all possible $^3J_{CH}$ coupling pathways and the minimum number of calculations required to produce an empirical relationship between $^3J_{CH}$ and Φ and Ψ.

	Rotation about Φ/Ψ Possible?		Number of substituent positions (n)			Number of substituent combinations: X(H, Me, OH, NH_2, F), Y(CH_2, NH, O) and Z(CH, N)				Number of calculations[c]
Coupling pathway label[a]	**Φ**	**Ψ**	**X**	**Y**	**Z**	**X** (5^n)	**Y** (3^n)	**Z** (2^n)	**Final**	**Varying Φ and/or Ψ in 30° steps**
333	Y	Y	7	0	0	78125	1	1	78125	2.6×10^7
1e33	Y	N	4	0	1	625	1	2	1250	3.3×10^4
3i1i1i	N	Y	3	0	0	125	1	1	125	3.3×10^3
1i1i3	Y	Y	3	0	0	125	1	1	125	4.2×10^4
2e33	Y	Y	5	1	0	3125	3	1	9375	3.2×10^6
32e3	Y	Y	5	1	0	3125	3	1	9375	3.2×10^6
332e	Y	Y	5	1	0	3125	3	1	9375	3.2×10^6
2a33	Y	Y	5	1	0	3125	3	1	9375	3.2×10^6
2e2e3	Y	Y	3	2	0	125	9	1	1125	3.8×10^5
2i2i3	Y	N	3	2	0	125	9	1	1125	2.9×10^4
2a2a3	Y	N	3	2	0	125	9	1	1125	2.9×10^4
2a2e3	Y	Y	2	3	0	25	27	1	675	2.3×10^5
32e2e	Y	Y	3	2	0	125	9	1	1125	3.8×10^5
2e32e	Y	Y	3	2	0	125	9	1	1125	3.8×10^5
2a32e	Y	Y	3	2	0	125	9	1	1125	3.8×10^5
32i2i	N[b]	Y	5	0	0	3125	1	1	6250	1.6×10^5
32a2a	N	Y	3	2	0	125	9	1	1125	2.9×10^4
2ei2i3	Y	N	3	1	0	125	3	1	375	9.8×10^3
1e2e3	Y	N	2	1	1	25	3	2	150	3.9×10^3
1e32e	Y	Y	2	1	1	25	3	2	150	5.1×10^4
1e1i1i	N	N	0	0	1	1	1	2	2	4.0×10^0
2i1i2i	N	N	3	0	0	125	1	1	125	2.5×10^2
1e2e2e	Y	N	0	2	1	1	9	2	18	4.7×10^2
1e2i2i	N[b]	N	2	0	1	25	1	2	100	2.0×10^2
1e2a2a	N	N	0	2	1	1	9	2	18	3.6×10^1
2e2e2e	Y	Y	1	3	0	5	27	1	135	4.6×10^4
2a2e2e	Y	Y	0	4	0	1	81	1	81	2.7×10^4
2e2i2i	N[b]	Y	3	1	0	125	3	1	750	2.0×10^4
2e2a2a	N	Y	1	3	0	5	27	1	135	3.5×10^3
2i2i2e	Y	N	3	1	0	125	3	1	375	9.8×10^3
2a2a2a	N	N	2	2	0	25	9	1	225	4.5×10^2
2a2a2a	N[b]	N	0	4	0	1	81	1	81	1.6×10^2
1ei1i2i	N	N	1	1	0	5	3	1	15	3.0×10^1
1i1i2e	Y	N	1	1	0	5	3	1	15	3.9×10^2
2e1i1i	N	Y	1	1	0	5	3	1	15	3.9×10^2
2a1i1i	N	N	0	2	0	1	9	1	9	1.8×10^1
								Total	**134699**	**4.1×10^7**

[a] Corresponding to the labelling system in Figure III.3. [b] The coupled proton in 32i2i, 1e2i2i, 2e2i2i and 2a2a2a coupling pathways may be *cis* or *trans* to the coupled carbon, resulting in 2 isomers. [c] Combined number of DFT geometry optimisations and NMR calculations for each combination of dihedral angles.

However, this set of fragments contains many combinations of both coupling pathway and substituents which are structurally rare, or which do not provide 3D structural information. Therefore, subsets of fragments (Figure IV.1, page 74) were examined in turn, steadily introducing complexity. In this chapter, initial studies (section III.2) were performed by selecting

66 fragments that have only proton and *sp*^{3} carbons in the coupling pathway and carbon substituents (Figure III.4, Fragment Index: HC001-066). This subset of fragments is commonly represented in branched sections of organic molecules. In addition, the generalised equations for $^3J_{CH}$ previously included some of these fragments[7,10], therefore the separate effects of a single α-substituent and β/γ-substituents could be confirmed and new interactions explored. In Chapter IV more complex fragments, such as coupling pathways including *sp*^{2}-hybridised atoms in internal/external vinyl groups (section IV.2) and carbonyl carbons (section IV.3) are explored. The presence of heteroatoms in the β-position of the coupling pathway was also a motif of interest and the interaction between this coupling pathway variation and substituent effects is explored in section IV.4.

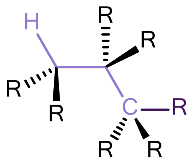


Figure III.4 Structure of the HCCC fragment, fragment index HC001-HC066 for R = H or Me.

III.1.2. The equation

A key consideration in the production of empirical relationships is the choice of equation; the equation must comprise terms capable of reproducing the observed behaviour. In the case of the Karplus equation ($^3J_{HH}$ as a function of Φ) the observed behaviour required an equation with maxima at Φ of 0 and ±180° and minima at ±90°, such as Equation I.13, where the values of the constants A, B and C are determined by regression analysis to predict $^3J_{HH}$ with the best fit to the measured/calculated scalar coupling constants. This work aims to incorporate the effects of α-, β- and γ-substituents into a single equation, therefore Equation I.19 which accounts for a single α-substituent[7] and Equation I.20 which accounts for β- and γ-substituents[10] were compared (Appendix 4, page 187). They are both contained within the Fourier series (Equation III.1) which van Beuzekom *et al.* used as a starting point to develop the α-substituent equation by eliminating unnecessary terms.

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{n}\sum_{j=0}^{n} C_{ij}\cos(i\Phi)\cos(j\Psi) + \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\sin(i\Phi)\sin(j\Psi) + \sum_{i=0}^{n}\sum_{j=1}^{n} T_{ij}\cos(i\Phi)\sin(j\Psi) + \sum_{i=1}^{n}\sum_{j=0}^{n} U_{ij}\sin(i\Phi)\cos(j\Psi) \qquad \text{Equation III.1}$$

Fourier series relating $^3J_{CH}$ to the dihedral angle Φ between the coupled ^{1}H and ^{13}C nuclei and the dihedral angle Ψ between an α-substituent and the nucleus the ^{1}H is attached.[7]

When considering molecules with a single α-substituent, van Beuzekom *et al.* noted that the mixed cosine and sine terms were not required due to the symmetry of the fragments and that

parameterisation showed that the constants (S_{ij}) attributed to the sine-only terms were negligible. The final equation, Equation I.19, used n = 2 for the cosine-only terms, which provided an adequate fit to the computationally derived $^3J_{CH}$. However, the rearrangement of Equation I.20, which accounts for β- and γ-substituents, shows that mixed sin(iΦ) × cos(jΨ) are required due to the changing symmetry of the fragments. Hence in this study, to ensure that terms necessary to account for the presence of substituent interactions were included, a 25-term variant of Equation III.1 with n = 2 for all four combinations of sine and cosine terms was proposed as a starting point for fitting the computationally determined $^3J_{CH}$. The final version of Equation III.1 that was necessary to account for the interactions between the coupling pathway and substituent effects investigated is described in more detail in Section III.2.6.

An additional consideration for Equation III.1 is the definition of the dihedral angle Ψ. In Equation I.19, the Ψ is defined as γ-β-α-X, where X is a single electronegative α-substituent and the other two α-substituents are protons. However, this definition is precluded by the introduction of multiple α-substituents. Therefore, Ψ was chosen to maximise the symmetry of the α-group on rotation to minimise the number of fragments requiring calculation. For example, in fragments with three α-substituents with one substituent different (*e.g.* H) to the other two (*e.g.* Me) then the unique substituent was used to define Ψ (*e.g.* γ-β-α-H). In the figures in this thesis, the substituent defining Ψ is drawn at Ψ = 180° and highlighted in dark purple. In Appendix 4, the substituent which is used to define Ψ is indicated (substituent α_1) for each fragment in the structural identifier.

III.1.3. Workflow

The fragment-based approach chosen to investigate the effects of substituents and coupling pathway on $^3J_{CH}$ used DFT for both geometry optimisation and NMR calculation of $^3J_{CH}$ (Gaussian 09 software[78]), bespoke scripts written in the AWK programming language to generate input calculation files and extract $^3J_{CH}$ data to transfer to MATLAB R2016a[168] for the regression analysis of Equation III.1. The first step, as outlined in Figure III.5, was the generation of the input geometries either as a z-matrix (b) so that the geometric parameters could be modified directly to generate all structural combinations, or as a set of Cartesian coordinates (a) for automated generation of different substituent combinations.

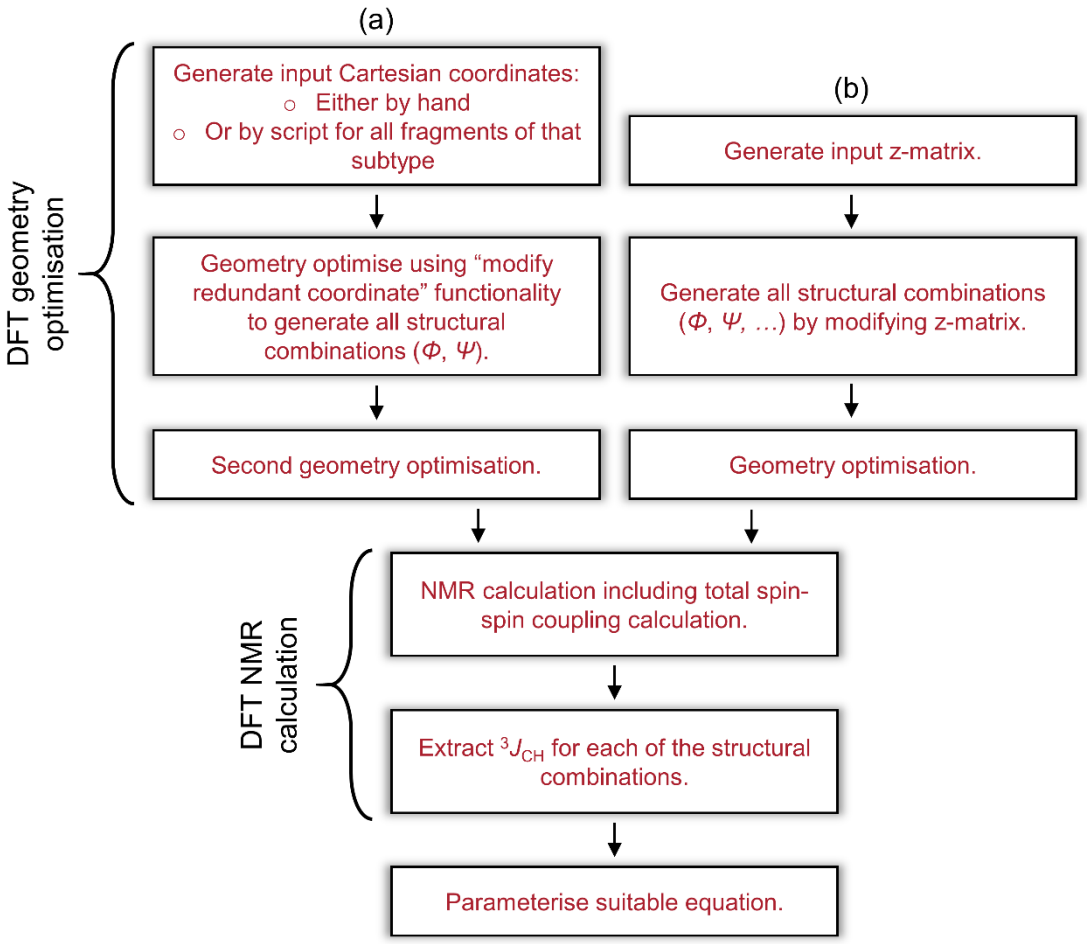


Figure III.5 Workflow for (a) Cartesian coordinate and (b) z-matrix molecular descriptions.

The input structures were then geometry optimised using DFT by either a one-step (for the structural combinations generated from z-matrix inputs) or a two-step (for molecules described by Cartesian coordinates) process. NMR calculations were performed with DFT to determine $^3J_{CH}$ for each structural combination using the same method described in Chapter II, including the empirically determined 6% correction. For each fragment a least squares regression analysis was performed using MATLAB R2016a to determine the relevant constants (C_{ij}, S_{ij}, T_{ij}, U_{ij}) for Equation III.1.

The full computational details for this process are provided in the experimental section (Chapter VIII). This resulted in a library of constants for Equation III.1 that were distinguished by the fragment used to generate them. The validation of these equations by comparison to experimentally measured $^3J_{CH}$ and calculated $^3J_{CH}$ is contained within Chapter V and Chapter VI respectively.

III.2. HCCC coupling pathway: saturated hydrocarbons

The relationship between $^3J_{CH}$, coupling pathway, substituent pattern and the dihedral angles Φ and Ψ was first explored for the coupling pathway containing sp^3 only carbons (HCCC). $^3J_{CH}$

was obtained as a function of these parameters (as outlined in the previous section, III.1.3) for a limited set of 66 fragments, which included all possible combinations of proton/methyl groups as substituents (Appendix 4).

III.2.1. Propane

The smallest saturated molecule examined was propane (Fragment Index HC001, Appendix 4), which consists of the sp^3 carbon fragment pathway and proton substituents only. The dihedral angles (Φ and Ψ) were varied from 0° to 360° in 30° steps to generate 169 structural variations, and therefore 169 DFT-calculated $^3J_{CH}$ values, Figure III.6.

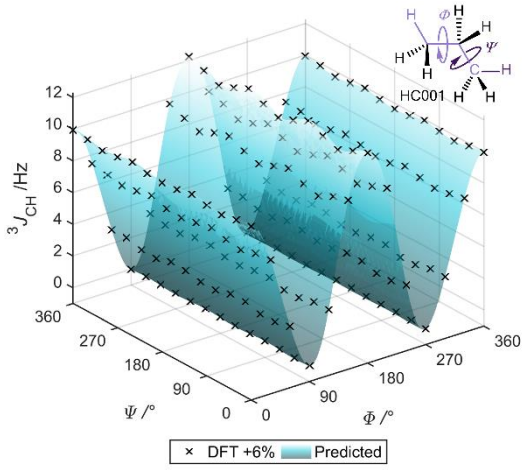


Figure III.6 DFT-calculated $^3J_{CH}$ for propane (HC001) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1 and is included for clarification of the 3D nature of the data.

The typical Φ-dependant relationship of vicinal scalar couplings was found in Figure III.6, with maxima at 0/360° and 180° where the coupled proton and carbon are eclipsed or anti respectively, and minima at 90° and 270° (for fixed Ψ). The second dihedral angle, Ψ, between an α-substituent and γ-carbon of the coupling pathway, instead showed maxima at 0°, 120° and 240° and minima at 60°, 180° and 300° (for fixed Φ). These maxima corresponded to geometries with an α-proton substituent eclipsed with the γ-carbon, whereas the minima were when an α-proton substituent was anti to the γ-carbon. However, it is clear from Figure III.6 that the effects of each dihedral angle on $^3J_{CH}$ are not independent. For example, when Φ is 180°, Ψ has the largest effect on $^3J_{CH}$, (±0.4 Hz) whereas for Φ = 0° and 360° this variation is even less (±0.1 Hz).

The DFT-calculated $^3J_{CH}$ values for propane ranged from 0.0-12.3 Hz and correspond to an average 5.6 Hz coupling under free rotation. This range is slightly larger than the 0.9-8.8 Hz range found by Wasylishen and Schaefer in 1972[6] for a HCC^{13}C fragment using the semi-empirical INDO (Intermediate Neglect of Differential Overlap) method to calculate the Fermi

contact (FC) contribution to $^3J_{CH}$ with changing Φ. The equation determined from their data (Equation III.2), was the first empirical equation for $^3J_{CH}$. However, experimental evidence suggesting that a Karplus-like relationship existed between Φ and $^3J_{CH}$ had been previously measured by Karabatso *et al.*[121] and with a greater range of Φ by Lemieux *et al.*[122]. In 1974, Wasylishen and Schaefer[169] reported the coupled ^{13}C spectrum for propane with an experimental $^3J_{CH}$ of 5.8 ±0.1 Hz, which is in good agreement with the averaged (under free rotation) 5.6 Hz value calculated by this work, and an improvement on the 4.3 Hz predicted by their Equation III.2.

$$^3J_{\mathrm{CH}} = 4.26 - 1.00\cos\Phi + 3.56\cos 2\Phi \qquad \text{Equation III.2}$$

Wasylishen and Schaefer equation, approximation of behaviour calculated for HCC^{13}C fragment.[6]

III.2.2. α-carbon substituents

The effect of α-substitution was explored by the addition of an α-methyl group (to give butane, HC002) and $^3J_{CH}$ were calculated by DFT as a function of Φ and Ψ, Figure III.7(a). The relationship between $^3J_{CH}$ and Φ for butane (HC002) was found to have a similar Karplus-like behaviour to propane. However, the effect of the second dihedral angle (Ψ) was more significant, up to ±0.9 Hz (Φ = 180°) in comparison to ±0.4 Hz for propane (HC001). The total range of $^3J_{CH}$ also decreased by 0.5 Hz to 0.0-11.8 Hz on addition of the methyl group. The minimum $^3J_{CH}$ remained at 0.0 ±0.1 Hz on addition of further α-methyl groups, isopentane (HC066) and 2,2-dimethyl butane (HC004) (Figure III.7(b/c)). The maximum $^3J_{CH}$ decreased, first to 11.4 Hz (two methyl groups) and then to 9.2 Hz (three methyl groups). This observed decrease in the maximum $^3J_{CH}$ may, in part, be due to the increasing steric bulk of the α-position and is reflected in the average bond angle between the α, β and γ carbons (θ_1), which increases from 115±2°, to 116±2° and finally to 117±1° as the number of methyl substituents increases. The bond length between the α and β-carbon ($r_{\alpha\beta}$), also increases as methyl substituents are added to the α-position from 1.538±0.009 Å, to 1.544±0.009 Å and finally to 1.554±0.009 Å.

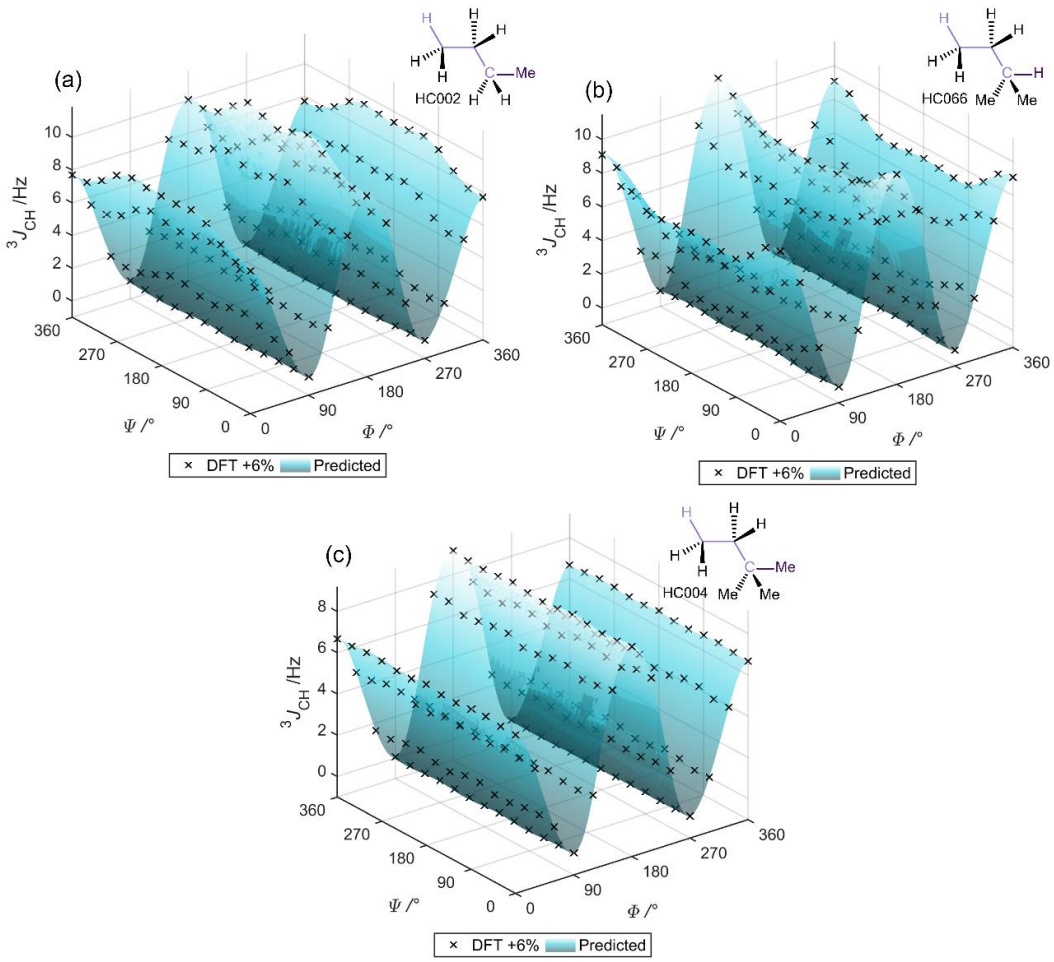


Figure III.7 DFT-calculated $^3J_{CH}$ for (a) butane (HC002), (b) isopentane (HC066) and (c) 2,2-dimethyl butane (HC004) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1 and are included for clarification.

The effect of Ψ on $^3J_{CH}$ calculated for propane agrees with the results of van Beuzekom *et al.*[7] (±0.4 Hz), however the semi-empirical INDO method that was used predicted a maximum $^3J_{CH}$ of 9.6 Hz for propane in comparison to the 12.3 Hz calculated by DFT. The maximum $^3J_{CH}$ calculated for butane was also lower than that calculated by DFT, giving 9.5 Hz compared to 11.8 Hz, and the effect of Ψ at Φ = 180° was lower, at ±0.4 Hz. However, the values of Φ and Ψ which led to the maximum $^3J_{CH}$ values agreed between the DFT and INDO calculated $^3J_{CH}$.

The relationship between $^3J_{CH}$ and Ψ was variable, following the maximum variation of ±0.9 Hz (at Φ = 180°) for butane (HC002) it increased to ±1.0 Hz for isopentane (HC066) and then decreased to ±0.1 Hz for 2,2-dimethyl butane (HC004). In propane the three maxima at Ψ = 0/360°, 120° and 240° (for fixed Φ) correspond to an α-proton and the γ-carbon in an eclipsed geometry. The maxima for butane (one α-methyl group) and isopentane (two α-methyl groups) are also where an α-proton and the γ-carbon are in an eclipsed geometry – there are two maxima at Ψ = 120° and 240° in butane, and one maximum at Ψ = 0° in isopentane. There is

also a smaller local maximum at Ψ = 180° in isopentane, where the α-proton and the γ-carbon are anti.

III.2.3. β-carbon substituents

The addition of methyl groups to the β-carbon gave isobutane (HC005 and HC006) and 2,2-dimethyl propane (HC007) which were used to examine the effect of β-substituents, Figure III.8. The addition of methyl groups to the β-position showed a decrease in the range of $^3J_{CH}$, while the minimum value remained at 0.0 ±0.1 Hz (the same as propane and the α-substituted examples) and the maximum $^3J_{CH}$ value decreased from 12.3 Hz (propane, HC001) to 11.0 Hz (isobutane, HC005/HC006) and finally to 9.8 Hz (2,2-dimethyl propane, HC007). The rotation of the α-protons (dependant on Ψ) shows the same effect on $^3J_{CH}$ as in propane – *i.e.*, maxima at Ψ = 0, 120 and 240° and minima at Ψ = 60 and 180° (for fixed Φ). The largest effect of Ψ on $^3J_{CH}$ was also when Φ is 180°, showing a ±0.4 Hz variation for isobutane (HC005/HC006) and ±0.3 Hz for 2,2-dimethyl propane (HC007).

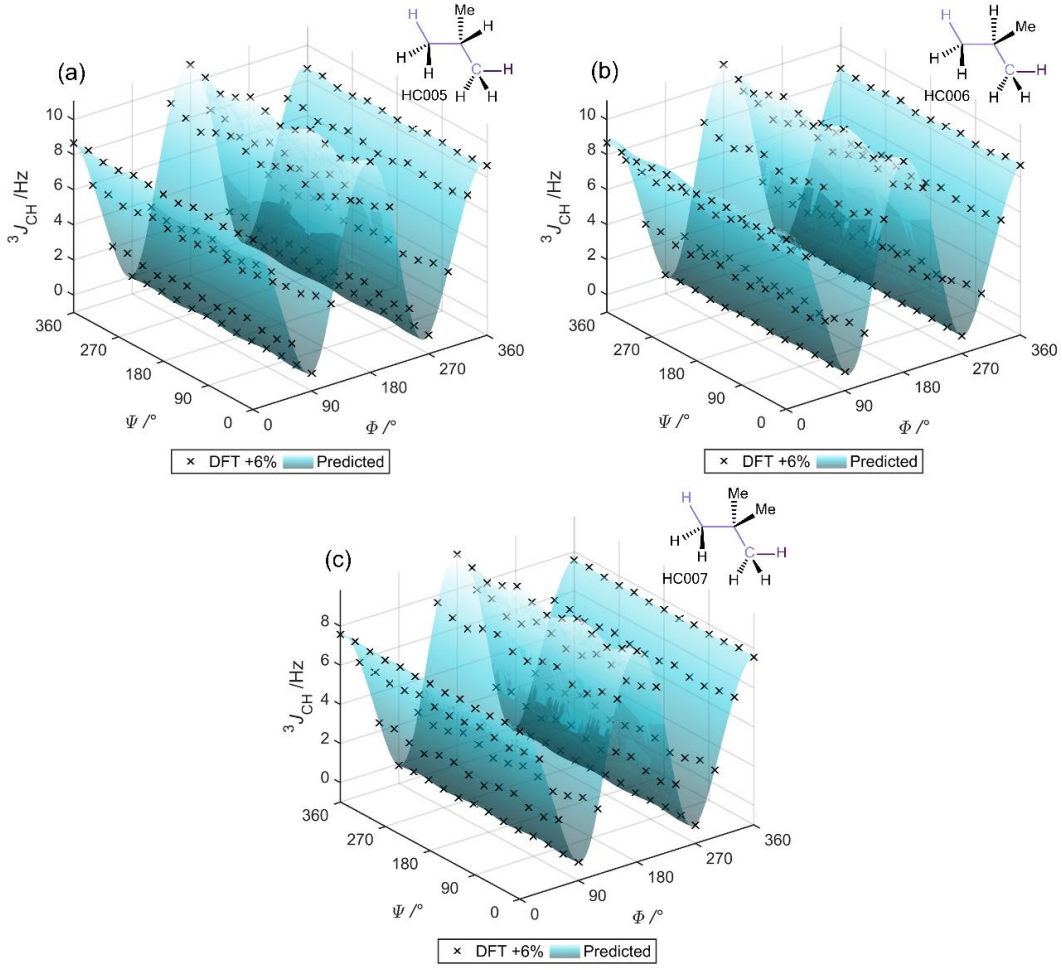


Figure III.8 DFT-calculated $^3J_{CH}$ for (a) isobutane (HC005), (b) isobutane (HC006) and (c) 2,2-dimethyl propane (HC007) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1 and are included for clarification.

The $^3J_{CH}$ calculated for the isobutane fragments, HC005 and HC006, showed a dependence on the dihedral angle Φ which was no longer symmetrical about 180°. This is shown more clearly in Figure III.9 where the Ψ-axis points directly into the page. For both HC005 and HC006, when the coupled γ-proton passes through an anti-relationship with the β-methyl substituent (Φ = 0-90° for HC005, Φ = 270-360° for HC006) the $^3J_{CH}$ coupling is smaller than the corresponding coupling for the mirrored value, where the coupled γ-proton passing through eclipsing the β-methyl group (Φ = 270-360° for HC005, Φ = 0-90° for HC006). However, between 90-270° for the Φ dihedral angle, if the coupled γ-proton is on the opposite face of the molecule to the β-methyl substituent (Φ = 90-180° for HC005, Φ = 180-270° for HC006) the coupling is larger than the corresponding dihedral where the γ-proton is on the same face as the β-methyl substituent (Φ = 180-270° for HC005, Φ = 90-180° for HC006).

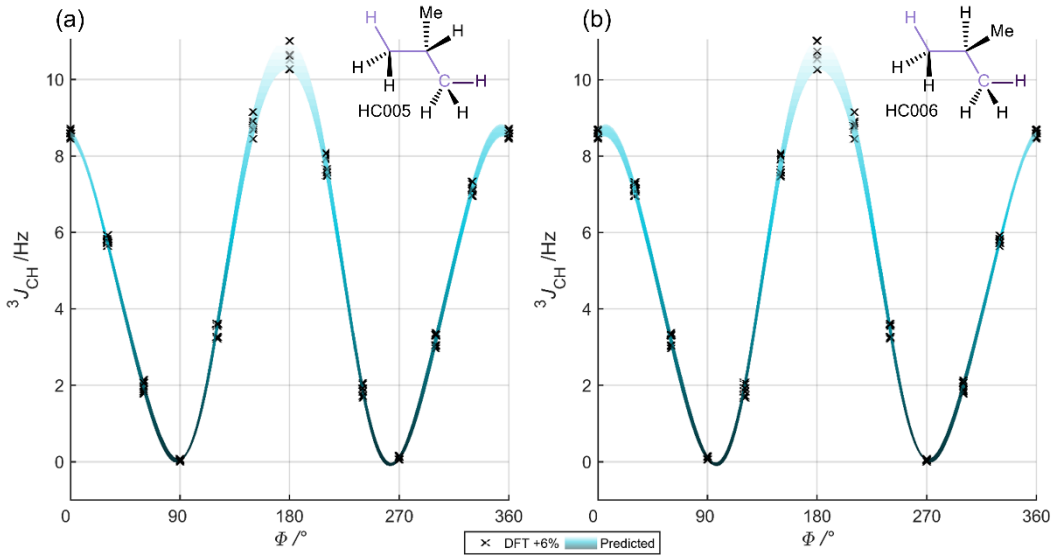


Figure III.9 DFT-calculated $^3J_{CH}$ for isobutane (a) HC005, (b) HC006 dependant on the dihedral angle Φ, black 'x' data points, (Ψ points into the page).

III.2.4. γ-carbon substituents

The addition of methyl groups to the γ-carbon gave butane (HC008 and HC009) and isopentane (HC010), for which $^3J_{CH}$ were calculated, Figure III.10. Following the addition of a single methyl substituent, these γ-substituted examples showed the most significant decrease (2.9 Hz) in the total $^3J_{CH}$ range, 0.0-9.4 Hz (Figure III.10(a/b), butane, HC008/HC009) in comparison to 0.0-12.4 Hz (propane, HC001). The addition of a second γ-methyl group (Figure III.10(c), isopentane, HC010) led to a further decrease of 0.8 Hz (0.1-8.6 Hz).

The $^3J_{CH}$ values calculated for these fragments containing γ-methyl groups showed that the dependence of $^3J_{CH}$ on Φ deviated from the typical Karplus-like relationship. The typical relationship is symmetrical about Φ = 180° and the maximum $^3J_{CH}$ values at 180° have a greater magnitude than those at Φ = 0/360° (Figure III.6). However, the addition of a single γ-

methyl group (butane, HC008/HC009) gave $^3J_{CH}$ values that approached the same maximum values, 9.1 ±0.3 Hz (180°) and 9.1 ±0.1 Hz (0/360°). For isopentane (HC010), which contains two γ-methyl groups, the maximum at Φ = 180° (7.0 ±0.1 Hz) was much less than that at Φ = 0/360° (8.4 ±0.1 Hz).

The Ψ dependence of $^3J_{CH}$ in molecules containing one γ-methyl group (butane, HC008/HC009) was found to give a variation of ±0.3 Hz at Φ = 180°, however the maximum variation was ±0.5 Hz at Φ = 210° for HC008 and Φ = 150° for HC009, where the γ-methyl group is at approximately 90° to the α-carbon. The isopentane molecule with two γ-methyl groups (HC010) showed a much smaller dependence on Ψ showing ±0.1 Hz at Φ = 180° with a maximum variation at Φ = 60/300° of ±0.2 Hz.

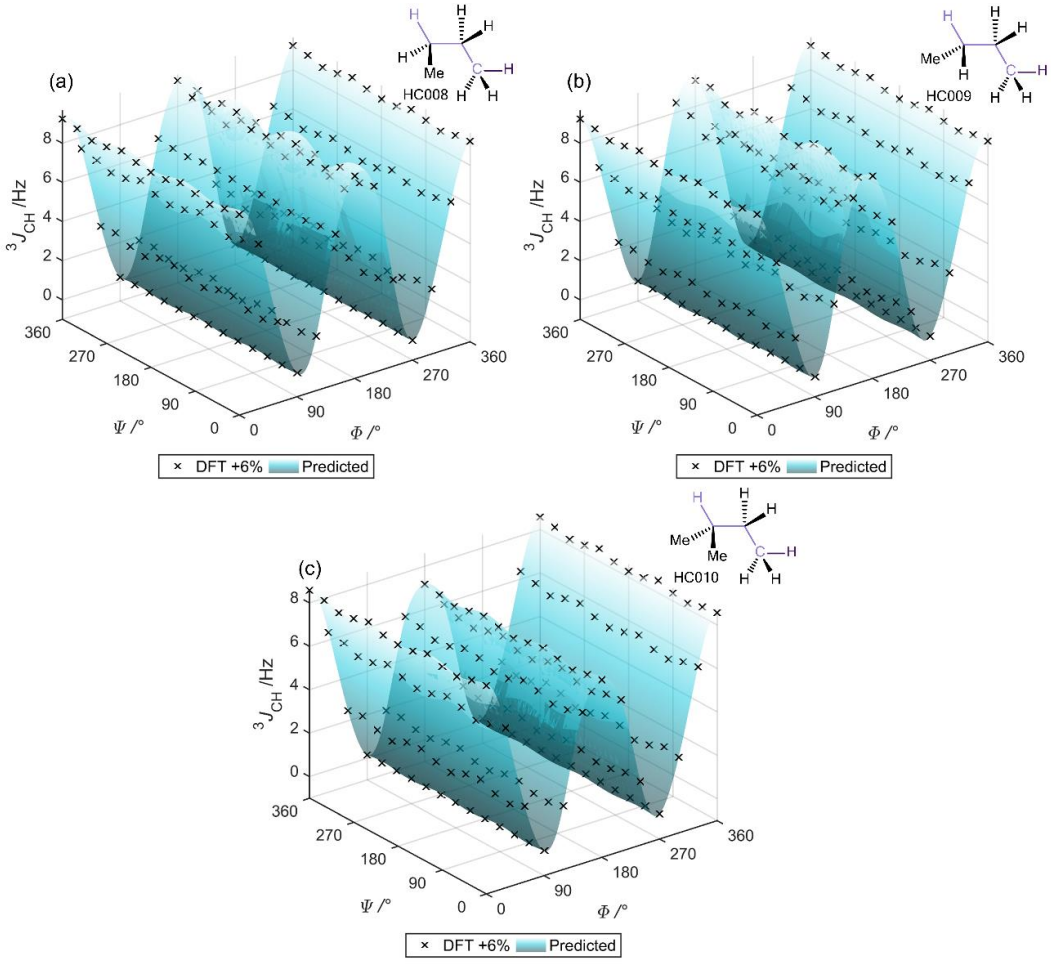


Figure III.10 DFT-calculated $^3J_{CH}$ for (a) butane (HC008), (b) butane (HC009) and (c) isopentane (HC010) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1 and are included for clarification.

The asymmetry about Φ = 180° in the $^3J_{CH}$ calculated for butane (HC008/HC009) is clearer in Figure III.11(a/b), with the $^3J_{CH}$ data viewed with the Ψ-axis pointing into the plane of the paper. For fragment HC008 for Φ = 0° to 90° the γ-methyl group passes from an eclipsed to an anti-

orientation to the β-protons. The $^3J_{CH}$ for this set of Φ dihedral angles are ~1.0 Hz larger than the mirrored $^3J_{CH}$ at Φ = 270-360° where the γ-methyl group changes from an eclipsed to anti-orientation with respect to the α-carbon. The values of $^3J_{CH}$ for Φ = 90-180°, where the γ-methyl group passes through an eclipsed orientation to the α-carbon, are smaller (1.0-1.5 Hz) than the mirrored values (Φ = 180-270°) where the γ-carbon is eclipsed with a β-proton. These same effects are present in fragment HC009 (Figure III.11(b)), however the γ-methyl group is on the opposite face of the molecule and therefore the corresponding Φ are reversed.

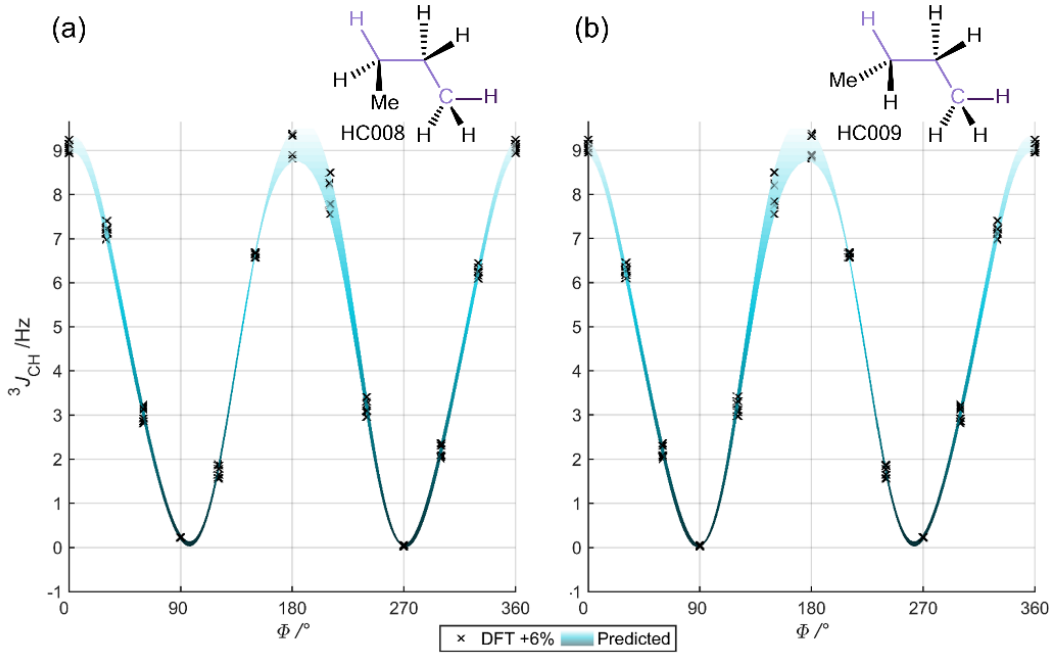


Figure III.11 DFT-calculated $^3J_{CH}$ for isobutane (a) butane (HC008) and (b) butane (HC009) dependant on the dihedral angle Φ, black 'x' data points, (Ψ points into the page).

III.2.5. Interaction between α, β and γ-substituents

This chapter has, so far, discussed ten HCCC fragments in detail, with a focus on confirming the basic relationship between $^3J_{CH}$ and the dihedral angles Φ and Ψ. These fragments initially contained no methyl substituents (propane, HC001), then the isolated effect of methyl substituents, attached only to the α-carbon (HC002/HC066/HC004), β-carbon (HC005/HC006/HC007) or γ-carbon (HC008/HC009/HC010) were explored. However, a key target of this work is the interaction between multiple substituents in various combinations of positions. Therefore, $^3J_{CH}$ were calculated for the remaining 56 fragments formed from combinations of 2-7 α, β and γ-methyl groups from the total 66 fragments (Appendix 4).

In Figure III.12 the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT is related to the number of α, β or γ-methyl groups from the different HCCC fragments. The first plot (Figure III.12(a)) shows the total range for fragments without γ-methyl groups. For the fragments also without β-methyl groups (solid line) small decreases in the total range with the

addition of 1-2 α-methyl groups were observed, while a much larger decrease was observed for the fragment with three α-methyl groups (as discussed in Section III.2.2). The introduction of β-methyl groups to fragment with α-methyl groups (dotted line, β = 1; dashed line β = 2) led to further decreases in the total range of $^3J_{CH}$ for the fragments. Generally, the addition of γ-methyl groups, γ = 1 (Figure III.12(b)) and γ = 2 (Figure III.12(b)) also decreased the total range of $^3J_{CH}$ for the fragments with the fully substituted fragment (2,2,3,3,4-pentamethylpentane, HC064, Appendix 4) giving the smallest range.

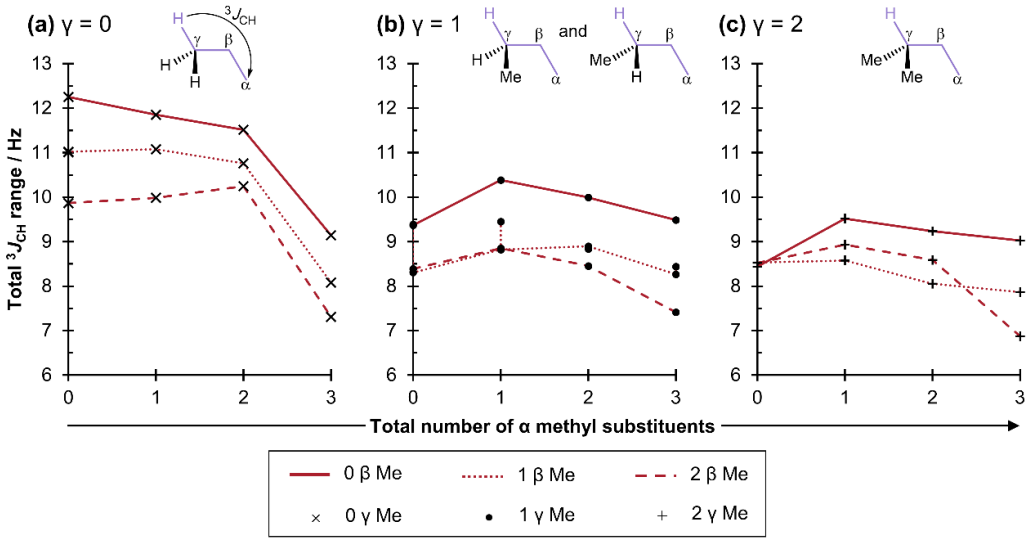


Figure III.12 Effect of the number of α, β and/or γ-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$).

However, the relationships between the total $^3J_{CH}$ range and the number of methyl substituents attached to α, β and/or γ-carbons in Figure III.12 show that the substituent effects on $^3J_{CH}$ are not independent. For example, three α-methyl groups led to an average decrease of 2.7 ±0.2 Hz relative to the total range of $^3J_{CH}$ for two α-methyl groups (in the absence of γ-methyl groups, Figure III.12(a)), while fragments containing 1-2 γ-methyl groups (Figure III.12(b/c)) showed a smaller average decrease (0.6 ±0.4 Hz).

The interactions between these substituents impact on the $^3J_{CH}$ value more than just the total range of coupling; the complex symmetry of the molecules formed by combining α, β and/or γ-methyl groups leads to an asymmetric relationship between $^3J_{CH}$ and Φ and/or Ψ about 180° in addition to further deviation from typical Karplus-like behaviour in terms of the exact positions (Φ and Ψ values) and relative magnitudes of the maxima and minima. Figure III.13 provides an example of the impact on $^3J_{CH}$ of the interaction between substituent effects and dihedral angle dependence for four fragments with a single methyl group attached to two positions (a-c) or all three positions (d). Isopentane (HC011, Figure III.13(a)) with α and β-methyl groups showed the largest maximum $^3J_{CH}$ (11.0 Hz) and the largest Ψ-effect of the four fragments

(±1.2 Hz, for fixed Φ). The $^3J_{CH}$ asymmetry with respect to Φ was similar to isobutane (HC006), which had a β-methyl group in the same position. However, the Ψ dependence, expected to be dominated by the α-substituent, deviates from that of butane (HC002); the relationship between $^3J_{CH}$ and Ψ was asymmetric about 180° for isopentane (HC011) whereas it was symmetric for butane (HC002), for Φ = 180° and 0/360°. This is highlighted by the relative magnitudes of the maxima in isopentane (HC011): the maximum at Ψ = 240°, where the β proton is eclipsed with the α-methyl, is larger than the maximum at Ψ = 120°, where the β-methyl and α-methyl groups are eclipsed (for Φ = 180°).

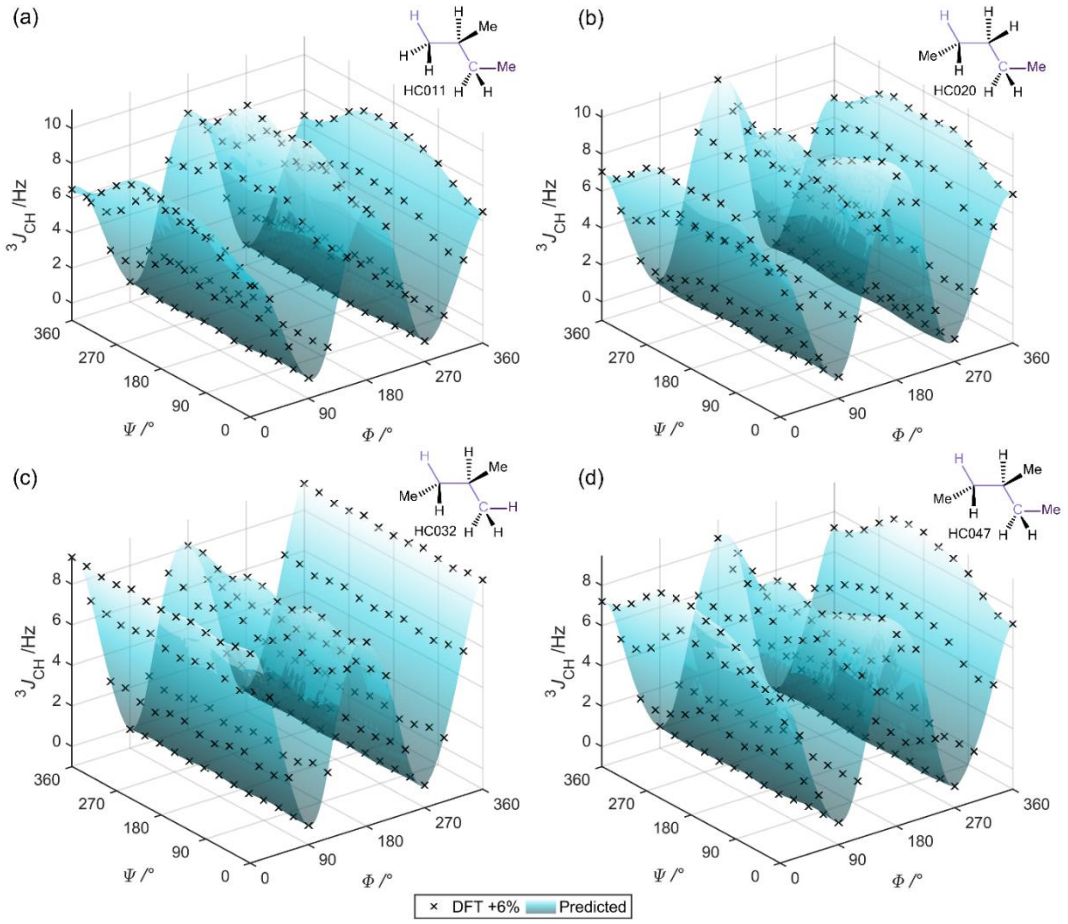


Figure III.13 DFT-calculated $^3J_{CH}$ for (a) isopentane (HC011), (b) pentane (HC020), (c) isopentane (HC032) and (d) 3-methyl pentane (HC047) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1 and are included for clarification.

The pentane fragment (HC020, Figure III.13(b)) with α and γ-methyl groups shows a reduced maximum $^3J_{CH}$ in comparison to isopentane (HC011), 10.4 Hz and 11.0 Hz respectively. The dependence of $^3J_{CH}$ on Φ is similar to isobutane (HC009), which has a single γ-methyl group and shows asymmetry about Φ = 180°. The maximum effect of Ψ was slightly reduced (±1.1 Hz) in comparison to isopentane (HC011) highlighting the difference between α-β and α-γ

methyl substituent interactions. The dependence of $^3J_{CH}$ on Ψ for Φ = 180° significantly deviated from the previously observed behaviour for a single α-methyl group (butane, HC002) for which there are two maxima at Ψ = 120/240° with approximately the same magnitude. Instead, for pentane (HC020) the magnitudes and positions of the two maxima differ, Ψ = 30° (10.4 Hz) and 210° (9.0 Hz). Although, the Ψ effect at Φ = 0/360 ° remained approximately the same as in the absence of a γ-methyl group.

The isopentane (HC032) fragment with β and γ-methyl groups, but without an α-methyl group shows a small maximum variation in $^3J_{CH}$ due to Ψ (±0.4 Hz), Figure III.13(c). However, the dependence of $^3J_{CH}$ on Φ significantly deviated from a typical Karplus-like behaviour; the dependence of $^3J_{CH}$ on Φ is asymmetric about Φ = 180°, additionally the average $^3J_{CH}$ at Φ = 180° is ~1.3 Hz less than the $^3J_{CH}$ at Φ = 0/360°.

Despite the presence of an additional α-methyl group there was not a significant variation in the maximum $^3J_{CH}$ (9.4 Hz) for 3-methyl pentane (HC047, Figure III.13(d)) compared to that of isopentane (HC032, 9.3 Hz). The maximum effect of Ψ of ±1.1 Hz (for fixed Φ) was similar to the two fragments in Figure III.13(a/b) which also contained an α-methyl group, isopentane (HC011) and pentane (HC020). The dependence of $^3J_{CH}$ on the dihedral angles Φ and Ψ is very complex for this fragment. For example, for Ψ = 0/360° the maximum at Φ = 180 ° is larger than that at Φ = 0/360°, however for Ψ = 180 ° this dependence is inverted and the maximum at Φ = 0/360° is larger than that at Φ = 180° and for both dihedral angles the relationship with $^3J_{CH}$ was asymmetric about 180°.

The four fragments discussed in detail in this section, with α, β and/or γ-methyl substituents, illustrate the importance of the interaction between substituents and the impact they have on the relationship between $^3J_{CH}$ and the dihedrals Φ and Ψ. The minimum $^3J_{CH}$ calculated was ~0 Hz for the 66 fragments (an average $^3J_{CH}$ of -0.01 ±0.07 Hz). The total range of $^3J_{CH}$ (Figure III.12) showed a decrease of ~0.5 Hz for each additional methyl substituent, however there was a significant variation in the relationship between $^3J_{CH}$ and the dihedrals Φ and Ψ depending on the number and position of the methyl groups. The dependence of $^3J_{CH}$ on Φ in these methyl substituted HCCC fragments is demonstrated in the data presented, for example the largest variation of ±6.1 Hz (Ψ = 240°) was observed for propane (HC001), while the smallest observed variation, ±3.4 Hz (Ψ = 240°) in the fully substituted fragment 2,2,3,3,4-pentamethylpentane (HC064), is also significant (Appendix 4). The variation due to Ψ is considerably smaller for the majority of the methyl substituted HCCC fragments. The largest variations of ±1.5 Hz are observed for HC016 (two α-methyl, two β-methyl) and HC022 (one α-methyl, two γ-methyl) however the maximum variation averaged 0.8 ± 0.4 Hz across the 66 fragments. Therefore, when considering the accuracy of experimental measurement of $^{n>1}J_{CH}$

(Table II.3, page 49), only Ψ-effects for fragments that show the greatest variation are significant, and only for comparison against $^3J_{CH}$ measured with highly accurate experimental techniques (such as the tilted EXSIDE 0.08/0.11 MAD/SD, or coupled ^{13}C).

III.2.6. Suitability of Equation III.1

In general, the more complex the substituent pattern, and therefore the more asymmetric the fragment, the more the dependence of $^3J_{CH}$ on Φ deviates from typical Karplus-like behaviour: exhibiting asymmetry about Φ = 180° and/or convergence/inversion of the relative magnitudes of the maxima at 0/360° and 180°. The effect of Ψ also varies with substituent pattern with changing relative magnitudes and position of maxima (for fixed Φ). Therefore for Equation III.1 (repeated below for clarity) to describe the complex relationship between $^3J_{CH}$, Φ and Ψ the maximum values of *i* and *j* need to be carefully considered for these more complex fragments.

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{n}\sum_{j=0}^{n} C_{ij}\cos(i\Phi)\cos(j\Psi) + \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\sin(i\Phi)\sin(j\Psi) + \sum_{i=0}^{n}\sum_{j=1}^{n} T_{ij}\cos(i\Phi)\sin(j\Psi) + \sum_{i=1}^{n}\sum_{j=0}^{n} U_{ij}\sin(i\Phi)\cos(j\Psi) \qquad \text{Equation III.1}$$

The comparison of the equations for α (Equation I.19) and β/γ (Equation I.20) substituents in Appendix 4 suggested Equation III.1 with n = 2 should be sufficient, resulting in a total of 25 terms. A linear least square regression was performed to determine the coefficients C_{ij}, S_{ij}, T_{ij} and U_{ij} for each of the 66 fragments (HCCC with methyl groups) and the root mean squared deviation (RMSD) between the DFT-calculated and empirically predicted $^3J_{CH}$ was calculated as an indicator of the performance. The 25-term variant Equation III.1 showed an average RMSD of 0.3 Hz across the 66 fragments, indicating a poor performance in predicting $^3J_{CH}$. Further, in Figure III.14(a), the residuals (the difference between the DFT-calculated and empirically predicted $^3J_{CH}$) of three selected fragments (i) propane (HC001), (ii) pentane (HC021) and (iii) 2,2,3,3,4-pentamethylpentane (HC064) showed systematic errors, in particular clustering of residuals for similar couplings. The residuals for propane (Figure III.14(a)i.), for example, are clustered around five different residual values (±0.04 Hz) for the 78 $^3J_{CH}$ that are <5 Hz.

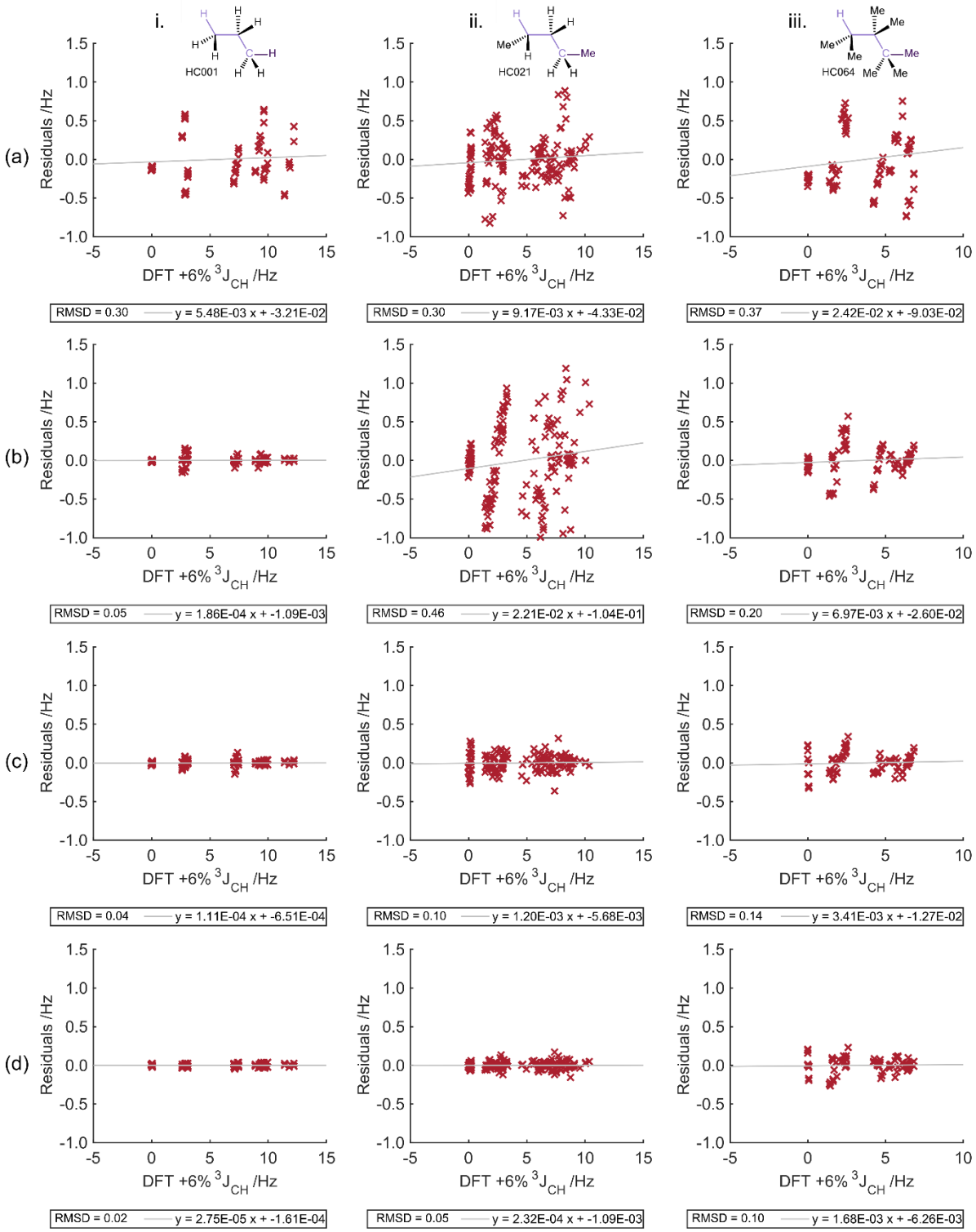


Figure III.14 Residuals between DFT-calculated and empirically predicted $^3J_{CH}$ for selected fragments i. propane (HC001), ii. pentane (HC021) and iii. 2,2,3,3,4-pentamethylpentane (HC064). $^3J_{CH}$ were predicted with the 25 term (a), 24 term (b), 56 term (c) and 75 term (d) variations of Equation III.1 outlined in Table III.2.

The poor performance of this 25-term variant of Equation III.1 in predicting $^3J_{CH}$ may arise from the lack of complex (α, β and γ) substituent patterns in Equation I.19 and Equation I.20 and hence cannot properly treat for mixed α, β and γ-substituents. Therefore, an equation suitable for the simplest fragment (propane, HC001) was first considered. As described in Section III.2.1, the dependence of $^3J_{CH}$ on the dihedral angles was symmetrical about 180° for both Φ and Ψ, this means that only the cos(iΦ)×cos(jΨ) terms in Equation III.1 should be required to model this behaviour. To achieve an RMSD of ≤0.1 Hz and eliminate the clustering for propane (HC001, Figure III.14(b)i) it was found that 24-term variant of Equation III.1 was required with a maximum i (i_{max}) = 5 and j_{max} = 3 for the C_{ij}, Table III.2(b). However, the RMSD for the pentane (HC021, Figure III.14(b)ii) fragment increased (to 0.5 Hz) on removal of terms including sine, as the equation cannot produce asymmetric behaviour about dihedral angles of 180°.

Table III.2 Variants of Equation III.1 investigated (a-k). The fragment index indicates the least substituted fragment with >0.1 Hz RMSD without the additional terms.

		(a)	(b)	(c)	(d)	(e)	(f)	(h)	(i)	(j)	(k)
Total number of terms		25	24	28	40	42	52	56	60	65	75
C_{ij}	i_{max}	2	5	5	5	5	5	5	5	5	6
	j_{max}	2	3	3	3	3	3	3	3	3	3
S_{ij}	i_{max}	2		2	2	2	2	2	2	2	4
	j_{max}	2		2	2	3	3	3	3	3	3
T_{ij}	i_{max}	2			2	2	2	4	4	4	4
	j_{max}	2			2	2	2	2	2	3	3
U_{ij}	i_{max}	2			2	2	4	4	5	5	5
	j_{max}	2			2	2	3	3	3	3	3
Fragment index			HC001	HC002	HC003	HC005	HC008	HC021	HC030	HC044	HC064

Fragments with methyl substituents required the inclusion of asymmetric (sine) terms in Equation III.1 to achieve RMSD ≤0.1Hz. Table III.2(c-k) shows the minimum i_{max}/j_{max} required to achieve an RMSD ≤0.1Hz for the fragment indicated. Figure III.14(d) shows that the residuals from $^3J_{CH}$ predicted by the 75-term variant of Equation III.1 for the selected fragments no longer has significant systematic errors.

The potential of this 75-term overfitting the 169 DFT-calculated $^3J_{CH}$ for each fragment was explored by calculating $^3J_{CH}$ by DFT for propane with Φ and Ψ from 0 to 360 in 5° steps (5,329 $^3J_{CH}$), Figure III.15(a). $^3J_{CH}$ were then predicted for this smaller step size using the coefficients (C_{ij}, S_{ij}, T_{ij}, U_{ij}) of Equation III.1 calculated by linear least squares regression of the 169 $^3J_{CH}$ dataset and gave a comparable RMSD, 0.03 Hz for the 5° step size and 0.02 Hz for the 30° step size.

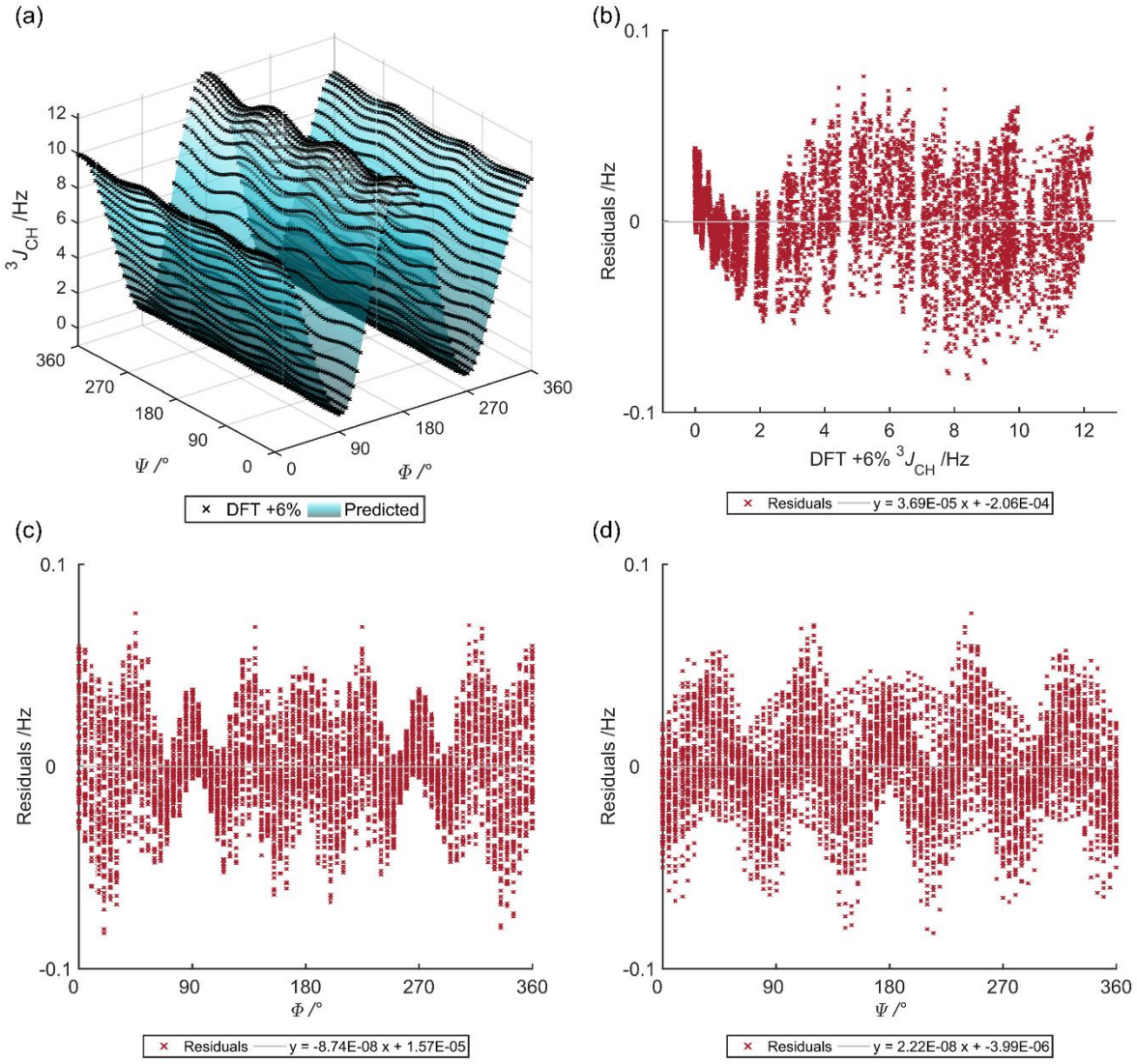


Figure III.15 (a) DFT-calculated $^3J_{CH}$ for propane (HC001) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1. Residuals between DFT-calculated and empirically predicted $^3J_{CH}$ using the 75-term variant of Equation III.1 were plotted against (b) DFT-calculated $^3J_{CH}$, (c) Φ and (d) Ψ.

Performing a least squares regression for the 75-term variant of Equation III.1 with the 5,329 $^3J_{CH}$ gave approximately the same coefficients as those from 169 $^3J_{CH}$ (average deviation of 0.0007 Hz) and an RMSD of 0.02 Hz. The primary difference between the fittings was a decrease (~80%) in the standard error in the coefficients reported by the fitting algorithm. The maximum residuals were larger for the smaller step size (approximately ±0.1 vs. ±0.05 Hz) Figure III.15(b), however these larger residuals were for geometric combinations that were not previously calculated with the 30° step size. Comparing the residuals to the dihedral angle Φ (Figure III.15 (c)) indicates that there is a small systematic error, however at least a cos(8Φ) term in Equation III.1 would be require to account for the symmetrical nature and five maxima between 0-180°. Additionally, comparing the residuals to the dihedral angle Ψ (Figure III.15

(d)) indicates that a cos(5Φ) term in Equation III.1 would be required to account for the symmetrical nature and five maxima between 0-180°. Given that the inclusion of more terms in Equation III.1 will increase the likelihood of overfitting the DFT-calculated $^3J_{CH}$ data it is not considered beneficial to increase the number of terms at this stage.

In summary, the prediction of $^3J_{CH}$ as a function of Φ and Ψ, with the 75-term variant of Equation III.1 (Table III.2(k)), showed good agreement with the $^3J_{CH}$ calculated by DFT for the 66 methyl substituted HCCC fragments (RMSD ≤0.1 Hz). Therefore, the 75-term equation was used throughout this thesis to fit DFT-calculated $^3J_{CH}$ data as a function of the dihedral angles Φ and Ψ for other fragments of interest, confirming an RMSD ≤0.1 Hz between the predicted and DFT-calculated $^3J_{CH}$ for all fragments.

Chapter IV. Empirical estimation of $^3J_{CH}$: incorporating complex fragments

This chapter focusses on the extension of the fragment-based approach to empirical $^3J_{CH}$ prediction for more complex fragments using the methods developed in Chapter III for HCCC fragments with proton/methyl substituents. These more complex fragments (Figure IV.1) included substituent and coupling pathway variation such as sp^2-hybridised atoms in internal/external vinyl groups (section IV.2) and carbonyl carbons (section IV.3). Heteroatoms (O/N) in the β-position of the coupling pathway was also a motif of interest and the interaction between this coupling pathway variation and substituent effects is explored in section IV.4. In total >500 fragments have been studied from the subtypes summarised in Figure IV.1.

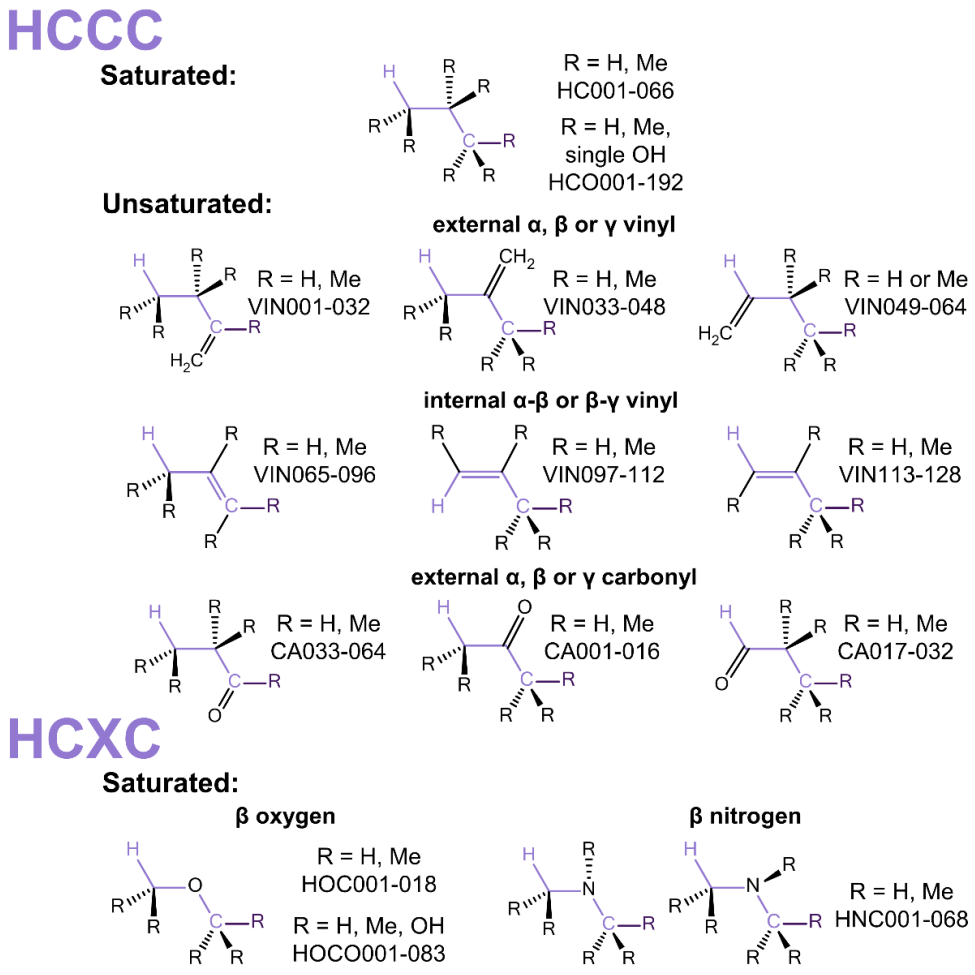


Figure IV.1 Subtypes of fragment coupling pathways targeted in this work and the substituent combinations explored (including the Fragment Index, Appendix 4).

IV.1. HCCC coupling pathway: additional substituent effects

IV.1.1. Carbonyl substituents

The $^3J_{CH}$ equations generated for HCCC sp^3 carbon-only fragments described in Chapter III used methyl groups as a proxy for any carbon substituent. Therefore, it was of interest to investigate the potential error in $^3J_{CH}$ predicted using this approximation, noting that substituent effects from atoms two bonds away from the coupling pathway are significant in the $^3J_{HH}$ coupling pathway[108]. Parella *et al*.[11] reported experimentally measured $^3J_{CH}$ for norbornane derivatives, which showed a 3.5 Hz effect on $^3J_{CH}$ on the introduction of an α-carbonyl substituent. Therefore, the variation between the effect of a carbonyl substituent and a methyl substituent was targeted.

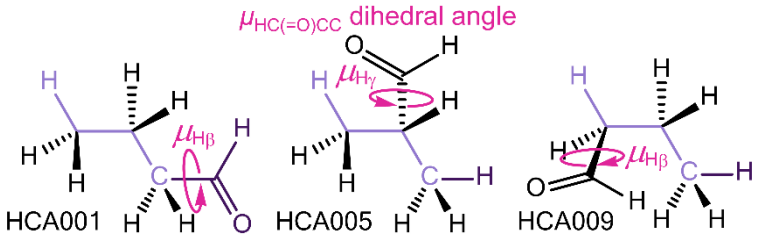


Figure IV.2 Definition of the dihedral angle, μ.

The carbonyl group introduces a third dihedral angle, μ, which is external to the coupling pathway, Figure IV.2. The effect of μ was included when exploring carbonyl substituted fragments by varying μ in 30° steps from 0-360° for each fragment in addition to Φ and Ψ. DFT was used to calculate the resulting 2,197 $^3J_{CH}$ for each fragment with a carbonyl substituent in the α (butanal, HCA001), β (isobutanal, HCA005 and HCA006) or γ (butanal, HCA009 and HCA010) position, Appendix 4. A fragment containing mixed methyl and carbonyl groups was also targeted: 3-methylpentanal (HCA172) with a γ-carbonyl substituent and α/β-methyl substituents. Initially, the 169 $^3J_{CH}$ calculated for μ = 180° for each of the fragments were used to fit Equation III.1 (RMSD ≤0.1 Hz), singly α/β/γ-substituted fragments are shown in Figure IV.3.

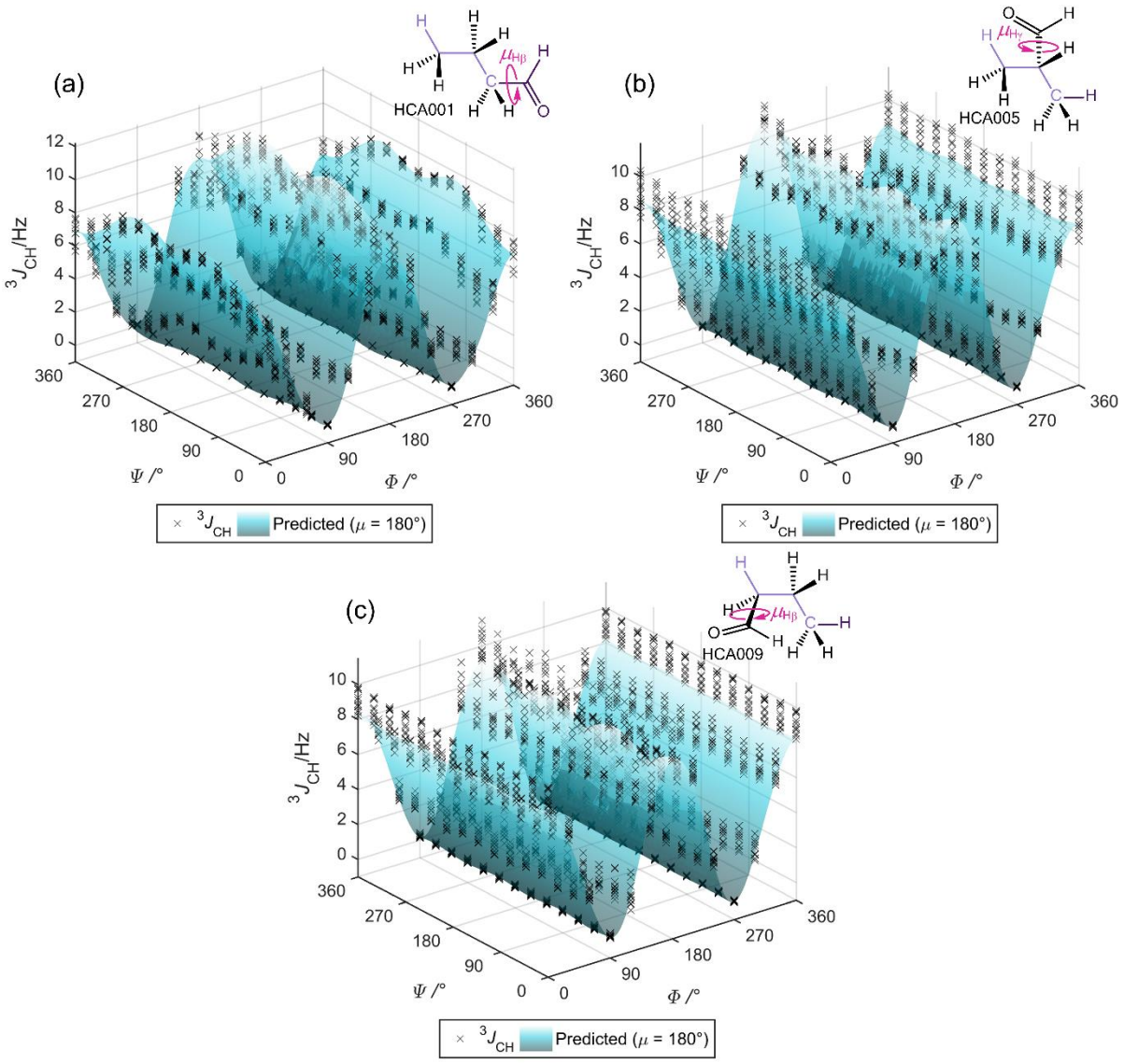


Figure IV.3 DFT-calculated $^3J_{CH}$ for (a) butanal (HCA001), (b) isobutanal (HCA005) and (c) butanal (HCA009) dependant on the dihedral angles *Φ, Ψ* and *μ*, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1 for μ = 180°.

However, despite Equation III.1 (parameterised with μ = 180°) performing well in predicting $^3J_{CH}$ where μ = 180° (RMSD ≤0.1 Hz), the prediction of $^3J_{CH}$ for all 2,197 geometric combinations of the six fragments (where μ = 0-360°) showed large RMSD (0.49-0.77 Hz), Table IV.1(a). Therefore, potential correction factors accounting for the effect of μ were explored.

Table IV.1 Summary of RMSD (Hz) from 2,197 DFT-calculated $^3J_{CH}$ for six carbonyl containing fragments compared to different methods of empirical prediction (a-f)

Fragment Index	HCA001	HCA005	HCA006	HCA009	HCA010	HCA172
(a) Equation III.1 (μ = 180°)	0.49	0.58	0.58	0.77	0.77	0.64
(b) Equation IV.1 (III.1+f'(μ))	0.44	0.45	0.45	0.48	0.48	0.48
(c) Equation IV.3 (III.1×f'(μ))	0.41	0.40	0.40	0.31	0.31	0.38
(d) Equation III.1 (average μ)	0.39	0.52	0.52	0.49	0.49	0.41
(e) Equation IV.4 f(Φ,Ψ,μ)	0.16	0.15	0.15	0.16	0.16	0.28
(f) Equation III.1 (methyl proxy)	0.55 (HC001)[a]	0.59 (HC005)[a]	0.59 (HC006)[a]	0.64 (HC008)[a]	0.61 (HC009)[a]	0.60 (HC048)[a]

[a] (fragment index for methyl proxy)

However, the introduction of a new geometric parameter, μ, requires the identification of a new equation, f(Φ,Ψ,μ). Initially, an additive correction to Equation III.1 (parameterised with $^3J_{CH}$ with μ = 180°) was proposed (Equation IV.1).

$$^3J_{\mathrm{CH}} = f(\Phi,\Psi,\mu) \approx f'(\Phi,\Psi) + f''(\mu) \qquad \text{Equation IV.1}$$

Where f'(Φ,Ψ) = Equation III.1 calculated with μ = 180° .

Therefore, the effect of μ was isolated from that of Φ and Ψ by subtracting the predicted $^3J_{CH}$ (μ = 180°) from DFT-calculated $^3J_{CH}$ (for all values of μ), matching the Φ and Ψ values, Figure IV.4i. A Fourier series, Equation IV.2, was used for this additive correction and a linear least squares regression[168] used to determine the C_h and S_h coefficients.

$$f''(\mu) = \sum_{h=0}^{3} C_h \cos(h\mu) + \sum_{h=1}^{3} S_h \sin(h\mu) \qquad \text{Equation IV.2}$$

This additive correction, Figure IV.4ii., improves the prediction of $^3J_{CH}$ (RMSD 0.44-0.48 Hz) in comparison to Equation III.1 (μ = 180°) (RMSD 0.49-0.77 Hz), however the accuracy is significantly less than the target RMSD of ≤0.1 Hz, Table IV.1(b).

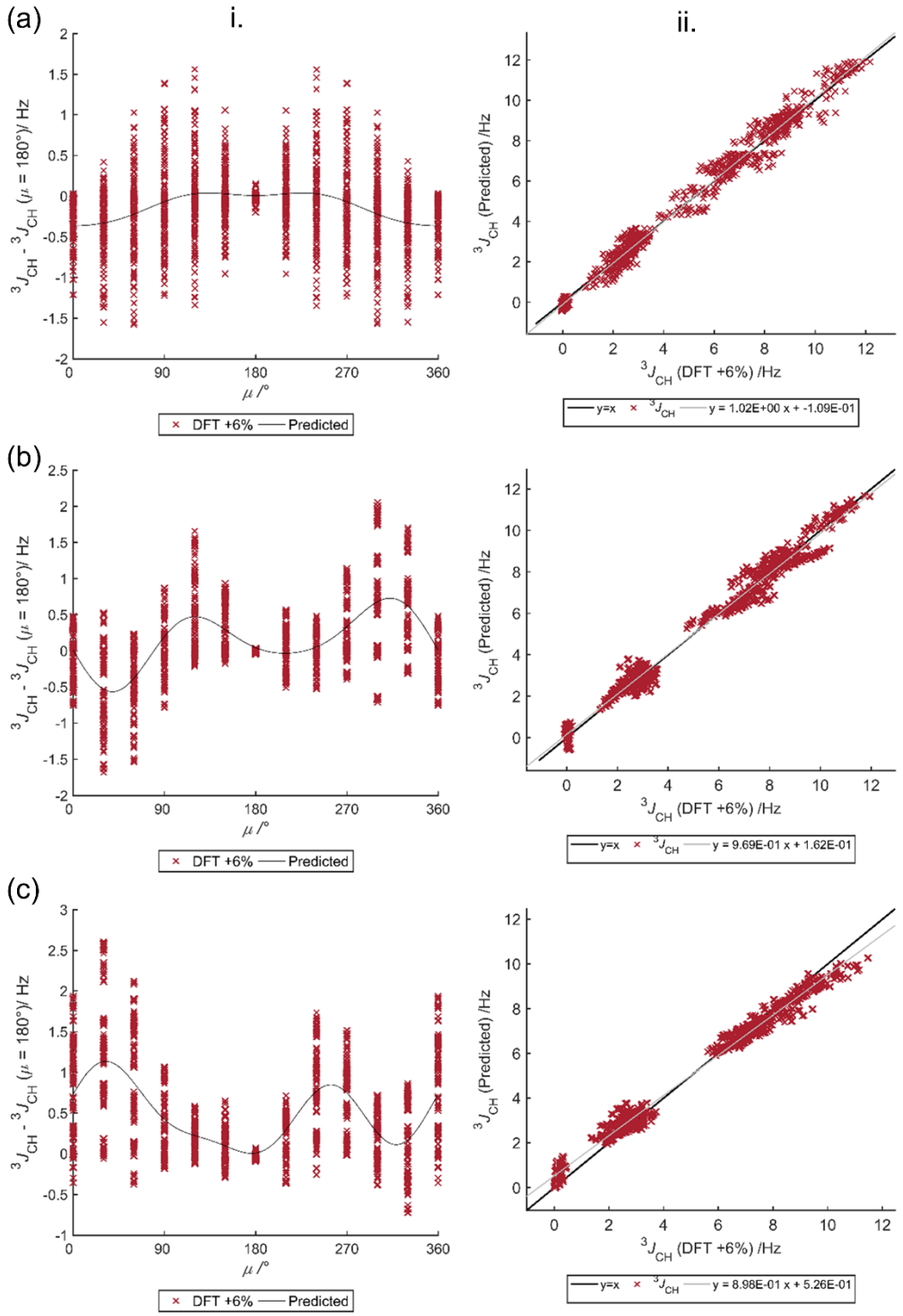


Figure IV.4 i. $^3J_{CH}$ predicted by Equation III.1 (μ = 180°) subtracted from DFT-calculated $^3J_{CH}$, red 'x' data points; f''(μ) (Equation IV.2), black line. ii. $^3J_{CH}$ predicted by Equation IV.1 compared to DFT-calculated $^3J_{CH}$. For fragments (a) butanal (HCA001), (b) isobutanal (HCA005) and (c) butanal (HCA009).

In the three plots in Figure IV.4i the range of deviations between the $^3J_{CH}$ predicted by Equation III.1 (μ = 180°) and the DFT-calculated $^3J_{CH}$ varies for fixed μ. For example, in butanal (HCA001, Figure IV.4(a)i.), when μ = 0° the range of deviation in $^3J_{CH}$ is approximately ±0.7

Hz, while for μ = 90° the deviation in $^3J_{CH}$ shows a range of approximately ±1.5 Hz. This highlights that the effects of Φ, Ψ and μ on $^3J_{CH}$ are not independent, and therefore a simple additive correction which applies the same correction for μ regardless of Φ and Ψ can only ever show a limited improvement. The small variation of DFT-calculated $^3J_{CH}$ at Φ/Ψ = 90° (where $^3J_{CH}$ ~0 Hz) compared to that at Φ/Ψ = 180° (where $^3J_{CH}$ ~10 Hz), shown in Figure IV.3 for each fragment, suggests that an improved approximation would be to assume that the interaction between Φ/Ψ and μ depends on the magnitude of $^3J_{CH}$. Therefore, a multiplicative correction was explored, Equation IV.3.

$$^3J_{\mathrm{CH}} = f(\Phi, \Psi, \mu) \approx f'(\Phi, \Psi) \times f''(\mu) \qquad \text{Equation IV.3}$$

Where f'(Φ,Ψ) = Equation III.1 (μ = 180°) and f''(μ) = Equation IV.2

For each of the six fragments, a linear least squares regression[168] was used to determine the C_h and S_h coefficients (Equation IV.3) from the 2,197 DFT-calculated $^3J_{CH}$. Figure IV.5i. shows the maximum and minimum possible $^3J_{CH}$ predicted by Equation IV.3 (black lines) as a function of μ, the majority of the DFT-calculated $^3J_{CH}$ are within this range. Figure IV.5ii. shows the predicted $^3J_{CH}$ compared to the DFT-calculated $^3J_{CH}$; it is an improvement on the additive correction of Equation IV.1, with the RMSD being reduced (0.31-0.41 Hz, Table IV.1(c)), however this correction is still larger than the RMSD of ≤0.1 Hz achieved for Φ/Ψ only effects.

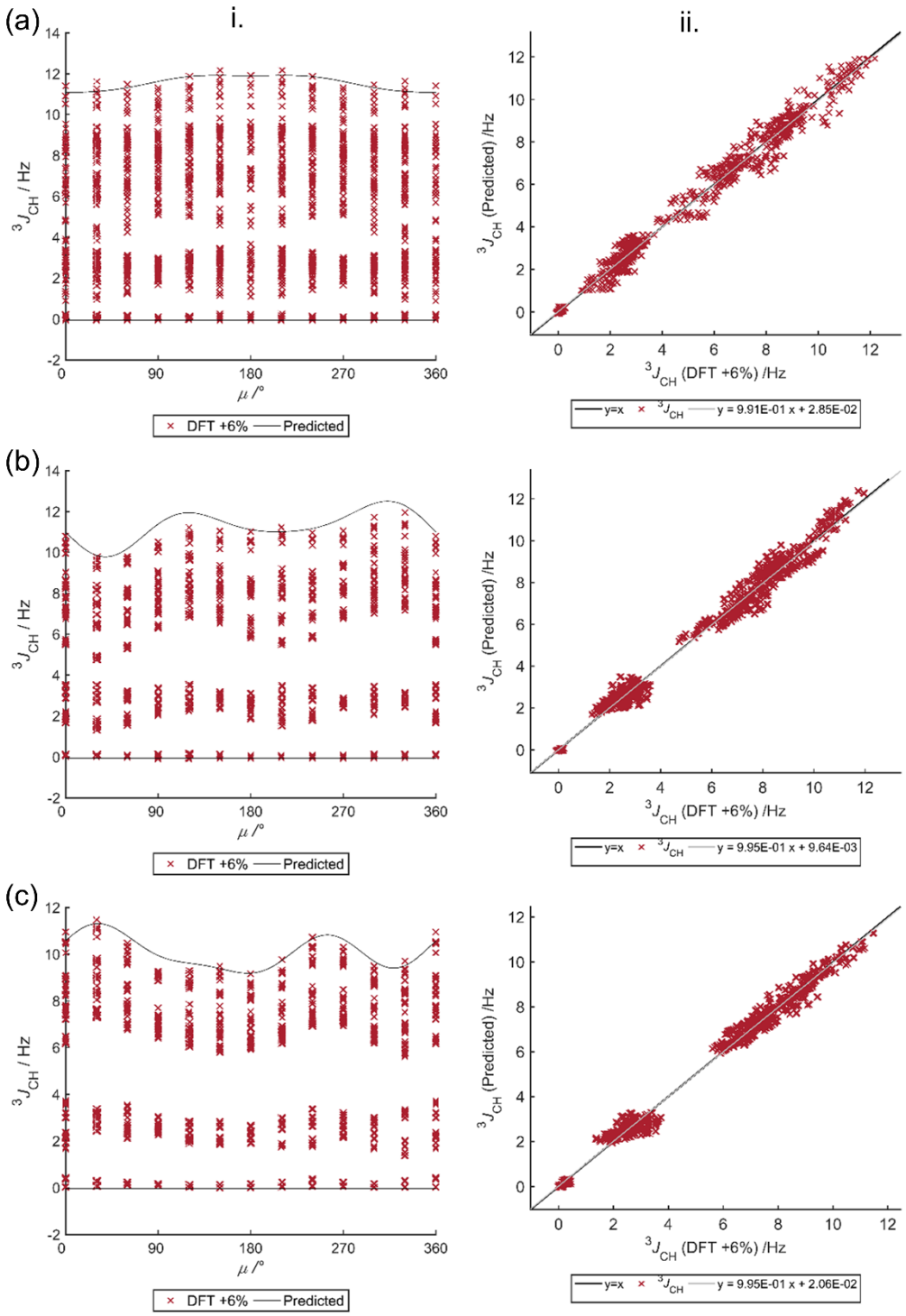


Figure IV.5 i. DFT-calculated $^3J_{CH}$, red 'x' data points; $^3J_{CH}$ predicted by Equation IV.3 (for maximum and minimum $^3J_{CH}$ predicted by Equation III.1), black lines. ii. $^3J_{CH}$ predicted by Equation IV.3 compared to DFT-calculated $^3J_{CH}$. For fragments (a) butanal (HCA001), (b) isobutanal (HCA005) and (c) butanal (HCA009).

Therefore an equation, f(Φ,Ψ,μ), that could reproduce the relationship between $^3J_{CH}$ and the interacting dihedral angles Φ, Ψ and μ was identified, Equation IV.4. This equation predicted

$^3J_{CH}$ with an RMSD ≤0.2 Hz (Table IV.1(e)) for the fragments containing only carbonyl and proton substituents.

$$
\begin{aligned}
{}^3J_{\mathrm{CH}} = & \sum_{i=0}^{6}\sum_{j=0}^{3}\sum_{h=0}^{3} C_{ijh}\cos(i\Phi)\cos(j\Psi)\cos(h\mu) \\
& + \sum_{i=0}^{6}\sum_{j=0}^{3}\sum_{h=0}^{0} C'_{ijh}\cos(i\Phi)\cos(j\Psi)\sin(h\mu) \\
& + \sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{h=0}^{3} S_{ijh}\sin(i\Phi)\sin(j\Psi)\cos(h\mu) \\
& + \sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{h=1}^{2} S'_{ijh}\sin(i\Phi)\sin(j\Psi)\sin(h\mu) \\
& + \sum_{i=0}^{4}\sum_{j=1}^{3}\sum_{h=0}^{0} T_{ijh}\cos(i\Phi)\sin(j\Psi)\cos(h\mu) \\
& + \sum_{i=0}^{4}\sum_{j=1}^{3}\sum_{h=1}^{3} T'_{ijh}\cos(i\Phi)\sin(j\Psi)\sin(h\mu) \\
& + \sum_{i=1}^{5}\sum_{j=0}^{3}\sum_{h=0}^{0} U_{ijh}\sin(i\Phi)\cos(j\Psi)\cos(h\mu) \\
& + \sum_{i=1}^{5}\sum_{j=0}^{3}\sum_{h=1}^{3} U'_{ijh}\sin(i\Phi)\cos(j\Psi)\sin(h\mu)
\end{aligned}
$$

Equation IV.4

As shown in Figure IV.6i., Equation IV.4, which is parameterised using DFT-calculated $^3J_{CH}$ from carbonyl containing fragments, can more accurately account for μ than the simpler additive (Equation IV.1, Figure IV.4ii.) or multiplicative (Equation IV.3, Figure IV.5ii.) corrections for α, β and γ carbonyl groups (HCA001, HCA005 and HCA009). Despite the accuracy of Equation IV.4 it requires the calculation of 2,197 $^3J_{CH}$ for all 192 combinations of a single carbonyl group with six proton/methyl groups, Appendix 4 (HCA001-HCA192). The marginal benefit in accuracy which would be obtained from these >800,000 DFT calculations is not considered worthwhile at this time.

The coefficients determined for Equation III.1 using methyl only fragments (HC001-HC066, as described in sections III.2.2-III.2.5) were used to predict $^3J_{CH}$ as a function of Φ and Ψ and compared to the DFT-calculated $^3J_{CH}$ from the carbonyl containing fragments as shown in Figure IV.6ii. The RMSD calculated from Equation III.1 with methyl proxies (0.55-0.60 Hz, Table IV.1(f)) show a similar accuracy to Equation III.1 (μ = 180°) parameterised with fragments containing a carbonyl group. Consequently, it appears that the benefits of introducing a μ-

correction for carbonyls are small, considering the number of calculations required. Therefore, using the coefficients determined for Equation III.1 for fragments with methyl-proxies (HC001-HC066) is proposed as a practical method.

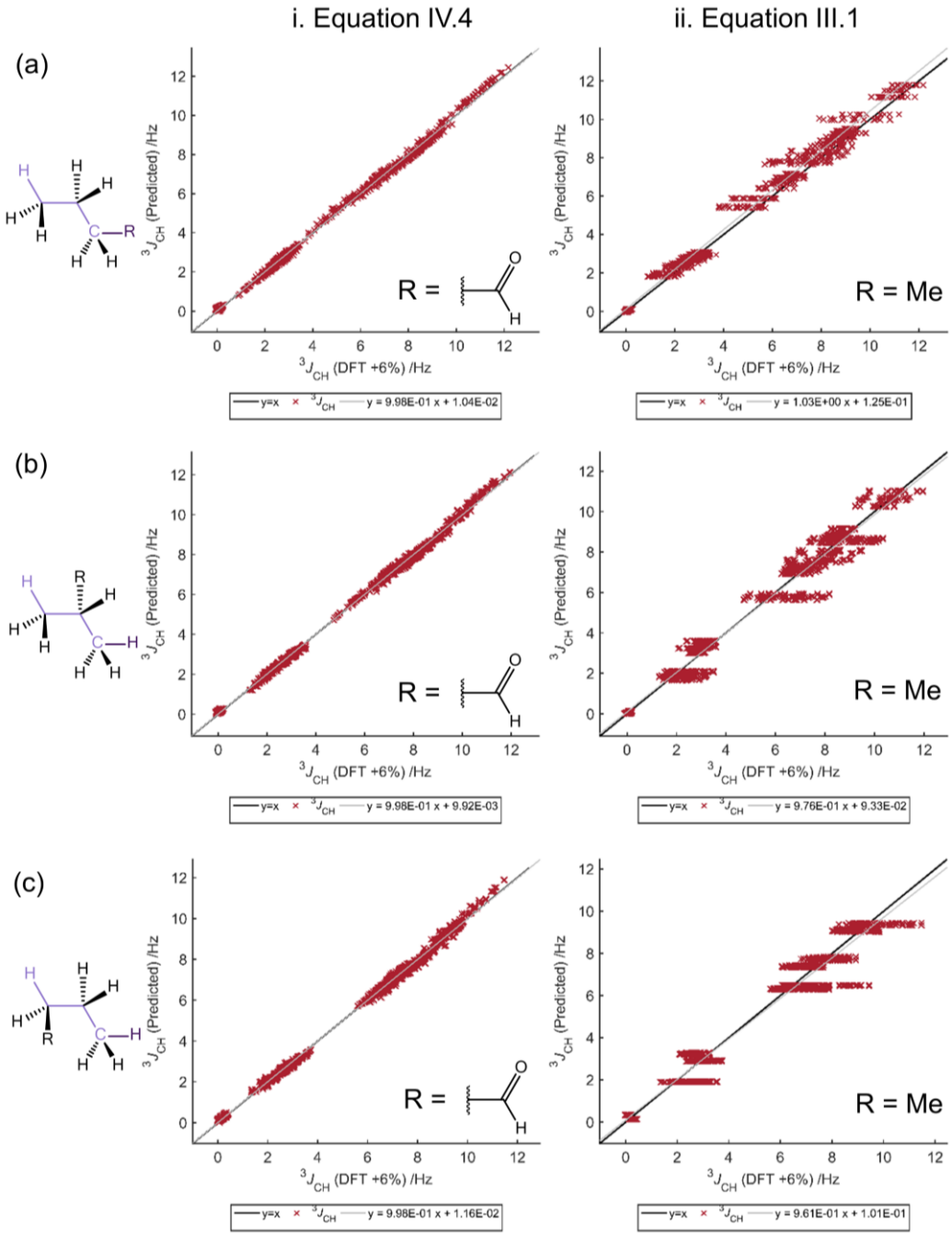


Figure IV.6 i. $^3J_{CH}$ predicted by Equation IV.4 compared to DFT-calculated $^3J_{CH}$. ii. $^3J_{CH}$ predicted by Equation III.1 (from fragments containing methyl proxy) compared to DFT-calculated $^3J_{CH}$. DFT-calculated $^3J_{CH}$ from (a) butanal (HCA001), (b) isobutanal (HCA005) and (c) butanal (HCA009).

IV.1.2. Oxygen substituents

The effect of oxygen substituents (OH groups) on DFT-calculated $^3J_{CH}$ was explored for the HCCC sp^3 carbon-only coupling pathway. The combination of 1-7 OH substituents with 0-6

proton/methyl groups forms 827 different fragments for this HCCC coupling pathway, requiring >250,000 DFT calculations to describe the relationship between $^3J_{CH}$, Φ and Ψ. Therefore, a smaller set of 192 fragments (Appendix 4) with single OH substituents were selected to investigate the differences between methyl and OH groups as substituents and any interactions between the substituents, therefore establishing the necessity of calculating the remaining fragments. The dihedral angle (μ) between the OH proton and the coupling pathway was set to 180° prior to geometry optimisation for all combinations of Φ and Ψ for each fragment. The DFT-calculated $^3J_{CH}$ for three fragments with α-OH (propan-1-ol, HCO001), β-OH (propan-2-ol, HCO005) or γ-OH (propan-1-ol, HCO009) are shown in Figure IV.7.

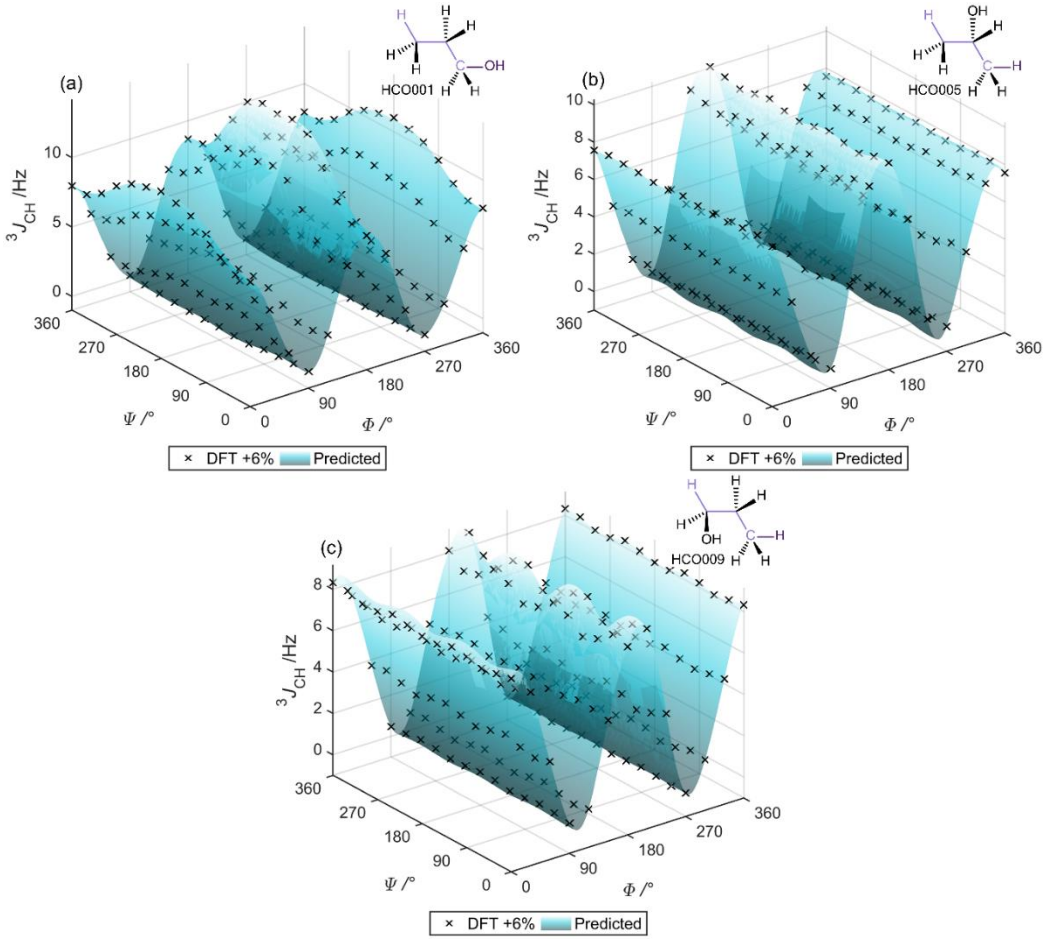


Figure IV.7 DFT-calculated $^3J_{CH}$ for (a) propan-1-ol (HCO001), (b) propan-2-ol (HCO005) and (c) propan-1-ol (HC0009) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1 and are included for clarification.

The $^3J_{CH}$ calculated for propan-1-ol (HCO001) with an α-OH substituent showed a typical Karplus-like relationship between $^3J_{CH}$ and Φ, Figure IV.7(a), however the range of $^3J_{CH}$ (-0.1-14.3 Hz) was larger than that of propane (HC001, 0.0-12.2 Hz) or butane (HC002, 0.0-11.8 Hz) with a single α-methyl group. The maximum effect of Ψ was also significantly increased (±2.7 Hz) for the α-OH group compared to that of either propane (±0.4 Hz) or butane (±0.9 Hz).

In propane, three maxima (for fixed Φ) were observed at Ψ = 0/360°, 120° and 240° which correspond to an α-proton and the γ-carbon in an eclipsed geometry. The two maxima for butane (one α-methyl group) were also with an α-proton and the γ-carbon in an eclipsed geometry (Ψ = 120° and 240°). However, for propan-1-ol (HCO001) when Φ = 0/360° there is only one maximum $^3J_{CH}$ at Ψ = 180°, but for Φ = 180° the values of $^3J_{CH}$ at Ψ = 150 and 210° are larger than that at Ψ = 180°.

The $^3J_{CH}$ calculated for propan-2-ol (HCO005), which contains a β-OH substituent, deviated from a symmetrical Karplus-like relationship around Φ = 180° (Figure IV.8(a)). As was observed for β-methyl substituted compounds, when the coupled γ-proton passes through an anti-relationship with the β-substituent (Φ = 0-90° for HCO005) the $^3J_{CH}$ coupling is smaller (~3.0 Hz for an OH group, ~1.0 Hz for a methyl group) than the corresponding coupling for the mirrored value, where the coupled γ-proton passes through eclipsing the β-OH group (Φ = 270-360°). However, between 90-270° for the Φ dihedral angle, if the coupled γ-proton is on the opposite face of the molecule to the β-substituent (Φ = 90-180° for HCO005) then the coupling is larger (~2.5 Hz for an OH group, ~1.0 Hz for a methyl group) than the corresponding dihedral where the γ-proton is on the same face as the β-substituent (Φ = 180-270° for HCO005). Unlike the α-substituted example, the range of $^3J_{CH}$ (0.0-10.1 Hz) was smaller for the β-OH substituted than the β-methyl substituted fragment, isobutane (HC005, 0.0-11.0 Hz). The maximum Ψ-effect for the β-OH substituted fragment was also lower (0.2 Hz) than the β-methyl substituted fragment, isobutane (HC005, 0.4 Hz).

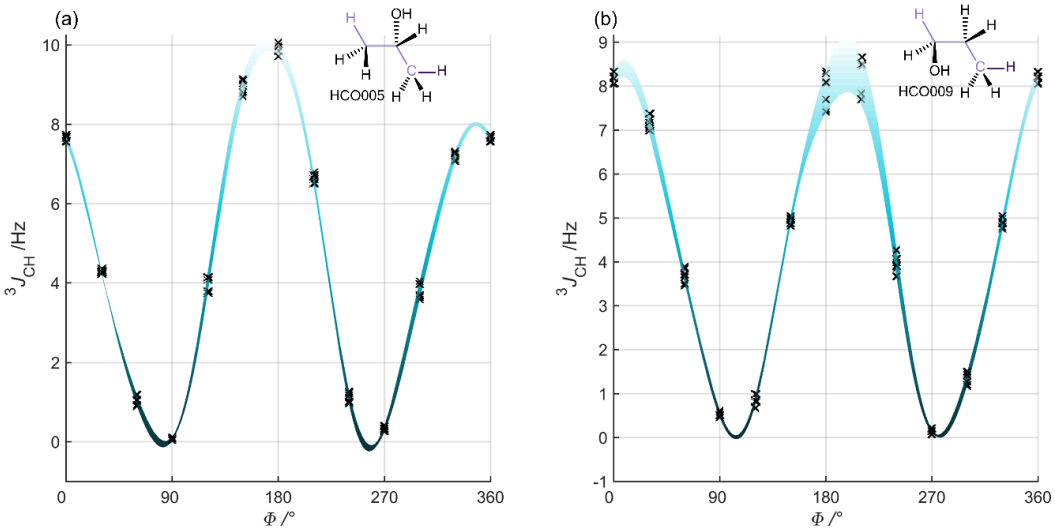


Figure IV.8 DFT-calculated $^3J_{CH}$ for (a) propan-2-ol (HCO005) and (b) propan-1-ol (HCO009) dependant on the dihedral angle Φ, black 'x' data points, (Ψ points into the page).

The range of $^3J_{CH}$ calculated for propan-1-ol (HCO009), which contains a γ-OH substituent, was smaller (0.1-8.7 Hz, Figure IV.7(c)) than the α-OH and β-OH substituted fragments and also smaller than the γ-methyl substituted fragment, butane (HC008, 0.0-9.4 Hz). The

maximum Ψ-effect for the γ-OH substituted fragment and the γ-methyl substituted fragment was approximately the same (0.5 Hz). The relationship between $^3J_{CH}$ and Φ deviated from a typical Karplus-like behaviour; Figure IV.8(b) shows the dependence of $^3J_{CH}$ on the dihedral angle Φ with the Ψ-axis pointing directly into the page. As was observed for γ-methyl substituted compounds, for Φ = 0° to 90° the γ-substituent passes from an eclipsed to an anti orientation to the β protons. The $^3J_{CH}$ for this set of Φ dihedral angles are ~2.0 Hz larger than the mirrored $^3J_{CH}$ at Φ = 270-360° where the γ-OH group changes from an eclipsed to anti orientation with respect to the α-carbon (this difference was ~1.0 Hz for a methyl group). The values of $^3J_{CH}$ for Φ = 90-180°, where the γ-OH group passes through an eclipsed orientation to the α-carbon, are smaller (~3.0 Hz) than the mirrored values (Φ = 180-270°) where the γ-carbon is eclipsed with a β proton.

$^3J_{CH}$ were also calculated by DFT for fragments with substituent patterns containing a mixture of 1-6 proton/methyl groups and a single OH group. The relationship between $^3J_{CH}$ and the dihedral angles (Φ and Ψ) increased in complexity (varying number of maxima, increased asymmetry and deviation from Karplus-like behaviour) due to the interactions between the OH and methyl groups in a similar manner as for methyl groups alone (section III.2.5). Figure IV.9 shows the effect on the total range of $^3J_{CH}$ from introducing methyl groups to fragments with an α-OH group, β-OH group or γ-OH group.

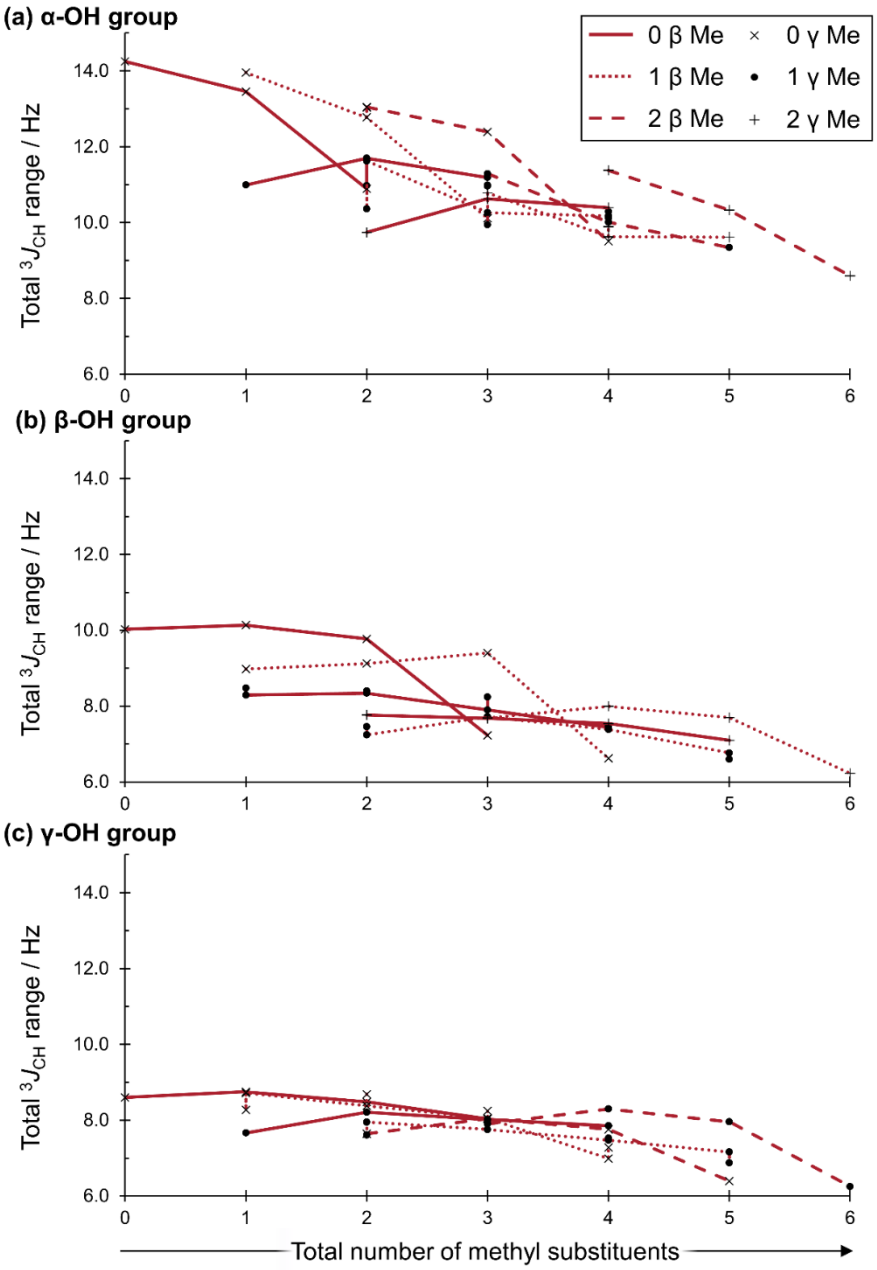


Figure IV.9 The effect of the total number of α, β and/or γ-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for an α (a), β (c) or γ (c) OH substituent.

The methyl substituents led to a larger decrease in the total range of $^3J_{CH}$ for fragments containing an α-OH (~0.8 Hz per methyl group), than for a β-OH group (~0.5 Hz per methyl group) or a γ-OH group (~0.3Hz per methyl group). For the methyl-only compounds (section III.2.5) each additional methyl group decreased the total range of $^3J_{CH}$ by ~0.5 Hz per methyl group.

In summary, the DFT-calculated $^3J_{CH}$ for each of these 192 fragments were used to fit 75-term variant of Equation III.1 (Table III.2(k)) and the predicted $^3J_{CH}$ showed good agreement (RMSD ≤0.1 Hz, Appendix 4). The differing impact of single OH substituents on $^3J_{CH}$ (depending on

their position and interaction with methyl substituents) suggests that calculating $^3J_{CH}$ for the remaining 635 fragments would be necessary to accurately describe the relationship between $^3J_{CH}$, Φ and Ψ for multiple OH substituents. At present these calculations have not been completed as variations in the coupling pathway (introducing sp^2 carbons) and the effect on the relationship between $^3J_{CH}$, Φ and/or Ψ were prioritised due to their prevalence in small organic molecules.

IV.2. HCCC coupling pathway: unsaturated hydrocarbons

The relationship between $^3J_{CH}$, coupling pathway, substituent pattern and the dihedral angles Φ and Ψ was explored for coupling pathways containing a mixture of sp^2 and sp^3 carbons (HCCC), Figure IV.10. $^3J_{CH}$ were calculated by DFT as a function of the dihedral angles Φ and Ψ for 64 fragments containing double bounds external to the coupling pathway and all possible combinations of proton/methyl groups as substituents (VIN001-VIN064, Appendix 4). The 64 fragments containing internal double bonds (VIN065-VIN128, Appendix 4) had only one dihedral angle to vary, for the α-β vinyl this was the Φ dihedral angle between the coupled nuclei, while for the *cis*/*trans* isomers of the β-γ vinyl this was the Ψ dihedral between the γ-carbon and an α-substituent.

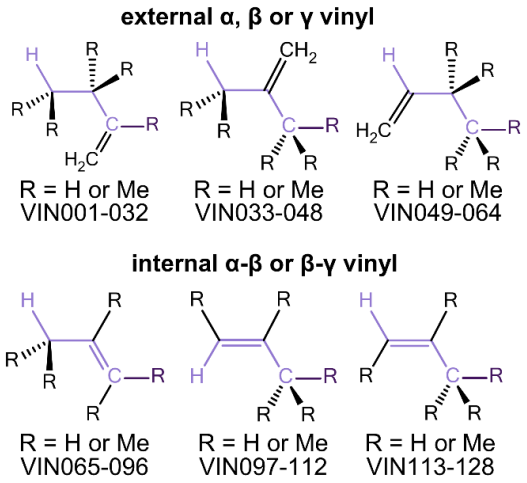


Figure IV.10 Subtypes of unsaturated HCCC fragments targeted including substituent combinations, including Fragment Index (Appendix 4).

IV.2.1. External α-vinyl group

The 1-butene (VIN001) fragment is the simplest example of a fragment with an external α-vinyl coupling pathway and only proton substituents. Figure IV.11 shows the relationship between $^3J_{CH}$ and Φ and Ψ for this fragment. The relationship between $^3J_{CH}$ and Φ shows a typical Karplus-like behaviour (for fixed Ψ) – with the maxima at Φ = 180° larger than those at Φ = 0/360° and symmetrical behaviour about Φ = 180°. The total range of $^3J_{CH}$ of -0.1-12.5 Hz is slightly larger than that for the saturated propane (HC001) and butane (HC002, α-methyl

group) described in section III.2. The $^3J_{CH}$, (Figure IV.11), show maxima at Ψ = 0/360° where the α-proton is eclipsed with the γ-carbon (for Φ = 0/360 and 180°). There is also a maximum for Ψ = 180° where the α-CH_2 group is eclipsed with the γ-carbon. The maximum Ψ-effect was ±1.3 Hz, larger than the ±0.4 Hz (propane, HC001) and ±0.9 Hz (butane, HC002) for the saturated fragments.

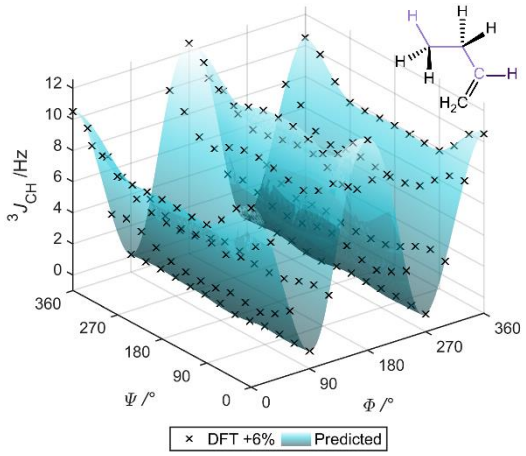


Figure IV.11 DFT-calculated $^3J_{CH}$ for 1-butene (VIN001) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

More complex substituent patterns for this coupling pathway were also of interest, therefore $^3J_{CH}$ were calculated for the remaining 31 fragments formed from combinations of 1-5 α, β and γ-methyl groups (VIN002-VIN032, Appendix 4). The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 32 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

In Figure IV.12 the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT is related to the number of α, β or γ-methyl groups from the fragments. The leftmost point for each line gives the total range for a fragment containing no α-methyl groups and the rightmost for a fragment containing one α-methyl group; for all combinations of β and γ-substituents the α-methyl group led to a decrease in the total range of $^3J_{CH}$, although the largest decreases (2.2-2.4 Hz) were for fragments with no γ-methyl groups (black 'x' data points). Across these 32 different fragments the minimum $^3J_{CH}$ remained approximately constant (-0.11±0.07 Hz). Each additional methyl led to a decrease of ~0.9 Hz in the total range of $^3J_{CH}$, larger than that observed for the saturated HCCC fragments (section III.2.5) of ~0.5 Hz per methyl group. In general, for these fragments containing α-vinyl groups for fragments without an α-methyl group the variation in $^3J_{CH}$ with Ψ were larger (±1.5 Hz) than the Ψ-effects for fragments with an α-methyl group (±0.9 Hz).

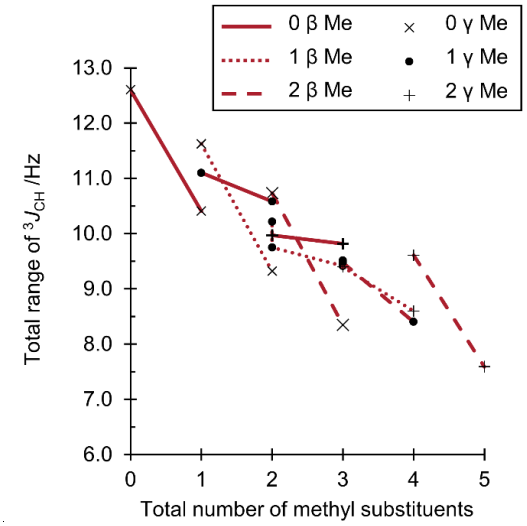


Figure IV.12 The effect of the total number of α, β and/or γ-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for an external α-vinyl group.

As was also observed for the saturated HCCC coupling pathway (section III.2.5), the interaction of substituents impacts on more than the total range of $^3J_{CH}$; the complex symmetry of the substituted molecules leads to an asymmetric relationship between $^3J_{CH}$ and Φ and/or Ψ about 180° in addition to further deviation from typical Karplus-like behaviour such as the positions (Φ and Ψ values) and relative magnitudes of the maxima and minima. For example, Figure IV.13 shows the DFT-calculated $^3J_{CH}$ for the β/γ-substituted fragment, 3,4-dimethyl-1-pentene (VIN029). In this fragment, when Ψ = 180° the $^3J_{CH}$ maximum at Φ = 180° is larger than that at 0/360° (closer to Karplus-like behaviour). However, when Ψ = 0/360° the maximum at the Φ = 0/360° is larger than that at 180°.

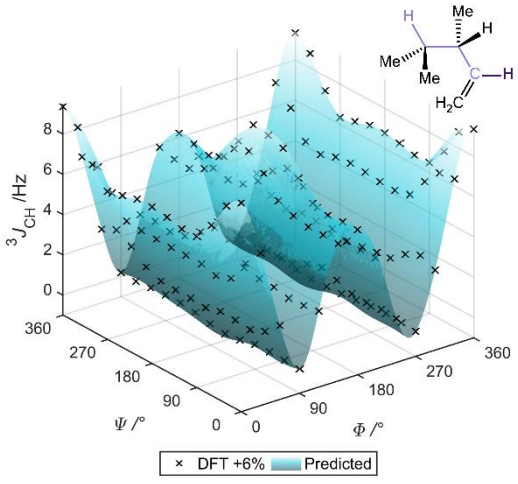


Figure IV.13 DFT-calculated $^3J_{CH}$ for 3,4-dimethyl-1-pentene (VIN029) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

IV.2.2. External β-vinyl group

The simplest of the fragments with an external β-vinyl coupling pathway was the fully protonated isobutene (VIN033) fragment. Figure IV.14 shows the relationship between $^3J_{CH}$ and Φ and Ψ for this fragment. $^3J_{CH}$ and Φ show a typical Karplus-like behaviour (for fixed Ψ), with the maxima at Φ = 180° larger than those at Φ = 0/360° and symmetrical behaviour about Φ = 180°. However, the minimum $^3J_{CH}$ of 1.1 Hz (at Φ = 90°) is larger than the 0 Hz typical of Karplus. The total range of $^3J_{CH}$ of 1.1-8.6 Hz was smaller than the external α-vinyl fragment 1-butene (VIN001) and isobutane (HC005/6, β-methyl group, described in section III.2). The $^3J_{CH}$, (Figure IV.14), show three maxima at Ψ = 0/360°, 120° and 240° where an α-proton is eclipsed with the γ-carbon (for fixed Φ). This effect is analogous to that observed for propane, which also contains three α-protons. The maximum Ψ-effect of ±0.4 Hz was also the same as propane (HC001).

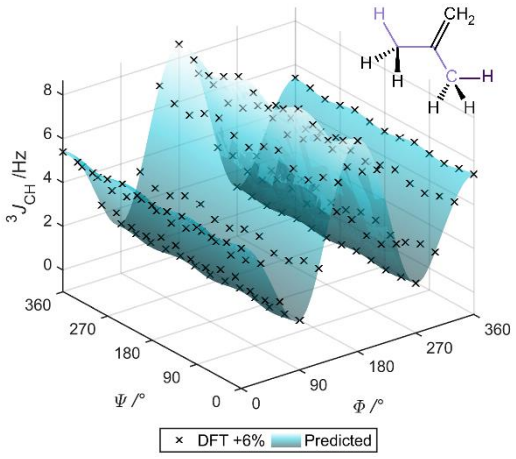


Figure IV.14 DFT-calculated $^3J_{CH}$ for isobutene (VIN033) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

The addition of 1-5 α and γ-methyl substituents to this coupling pathway resulted in 15 different fragments (VIN034-VIN048, Appendix 4). The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 16 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

In Figure IV.15 the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT is related to the number of α/γ-methyl groups in these fragments. The leftmost point for each line gives the total range for a fragment containing no α-methyl groups and the rightmost point for a fragment containing three α-methyl group. The addition of α-methyl groups generally had a small effect (±0.5 Hz) on the total range of $^3J_{CH}$, however for the VIN045 fragment with no γ-methyl groups the addition of the third α-methyl group (rightmost black 'x') causes a large ~2 Hz decrease. The addition of γ-methyl groups showed a larger effect on the total range of $^3J_{CH}$ than the α-methyl groups (except for VIN045). For example, for the fragments without α-methyl

groups the addition of one γ-methyl group led to a decrease in the total range of 2.2 Hz (leftmost 'x' vs. leftmost '●'), and a second γ-methyl group a further decrease of 0.4 Hz (leftmost '+').

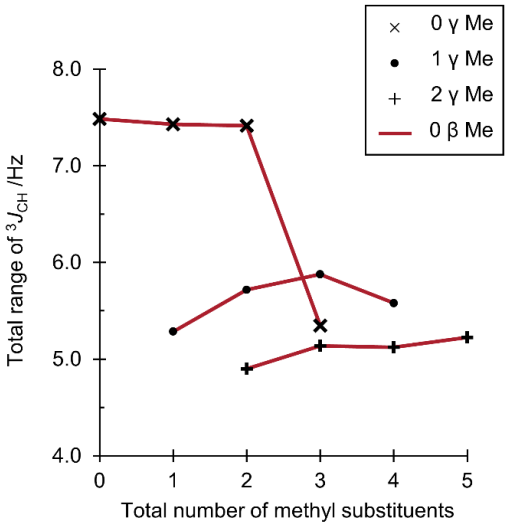


Figure IV.15 The effect of the total number of α and/or γ-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for an external β-vinyl group.

Across these 16 different fragments the minimum $^3J_{CH}$ remained approximately constant (0.9 ±0.1 Hz) and larger than 0 Hz. The main contribution to the changing total range of $^3J_{CH}$ was from the changing maximum $^3J_{CH}$: 8.62 Hz for the fully protonated fragment (VIN033) compared to 5.9 Hz for the fragment with 2 γ-methyl groups (VIN036). This deviates from the previously discussed coupling pathways where the smallest total range and the smallest maximum $^3J_{CH}$ have all been for the fully methylated fragments. For the external β-vinyl fragments each additional methyl led to a decrease of ~0.3 Hz in the total range of $^3J_{CH}$, smaller than that observed for the saturated HCCC fragments (section III.2.5) of ~0.5 Hz per methyl group.

IV.2.3. External γ-vinyl group

The relationship between $^3J_{CH}$ and Φ and Ψ for the external γ-vinyl coupling pathway is shown in Figure IV.16 for the 1-butene (VIN049) fragment. The relationship between $^3J_{CH}$ and Φ deviates from typical Karplus-like behaviour, with the maxima at Φ = 180° the same magnitude as those at Φ = 0/360°, although the $^3J_{CH}$ are symmetrical about Φ = 180°. The total range of $^3J_{CH}$ of 1.6-6.7 Hz is smaller than both the α- and β-vinyl fragments (VIN001 and VIN033 respectively). The $^3J_{CH}$, (Figure IV.16), show three maxima at Ψ = 0/360°, 120° and 240° where an α-proton is eclipsed with the γ-carbon (for fixed Φ). This effect is analogous to that observed for propane (HC001) and the β-vinyl fragment (VIN033), which also contain three α-protons. However, the maximum Ψ-effect was ±0.2 Hz, smaller than the ±0.4 Hz observed for propane (HC001) and the β-vinyl fragment (VIN033).

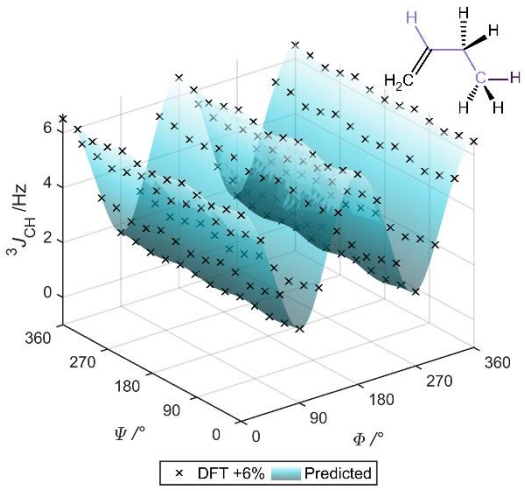


Figure IV.16 DFT-calculated $^3J_{CH}$ for 1-butene (VIN049) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

The addition of 1-5 α and β-methyl substituents to this coupling pathway resulted in 15 further fragments (VIN049-VIN064, Appendix 4). The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 16 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

In Figure IV.17 the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT is related to the number of α/β-methyl groups in the fragment. The leftmost point for each line gives the total range for a fragment containing no α-methyl groups and the rightmost point for a fragment containing three α-methyl group. The effect of adding methyl groups to this coupling pathway deviates from the previously discussed fragments as the total range of $^3J_{CH}$ increases (~0.2 Hz per methyl group) rather than decreases.

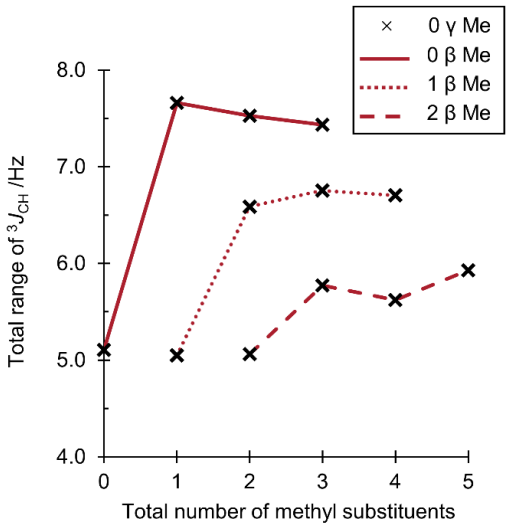


Figure IV.17 The effect of the total number of α and/or β-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for an external γ-vinyl group.

In the absence of α-methyl groups (the first 'x' of each line in Figure IV.17) the addition of β-methyl groups does not have a large effect on the total range of $^3J_{CH}$. However, the comparison of Figure IV.16 and Figure IV.18(a/b) shows the impact of the relative magnitude of the maxima at Φ = 0/360° and 180°, with each additional β-methyl group decreasing the maximum at Φ = 180°.

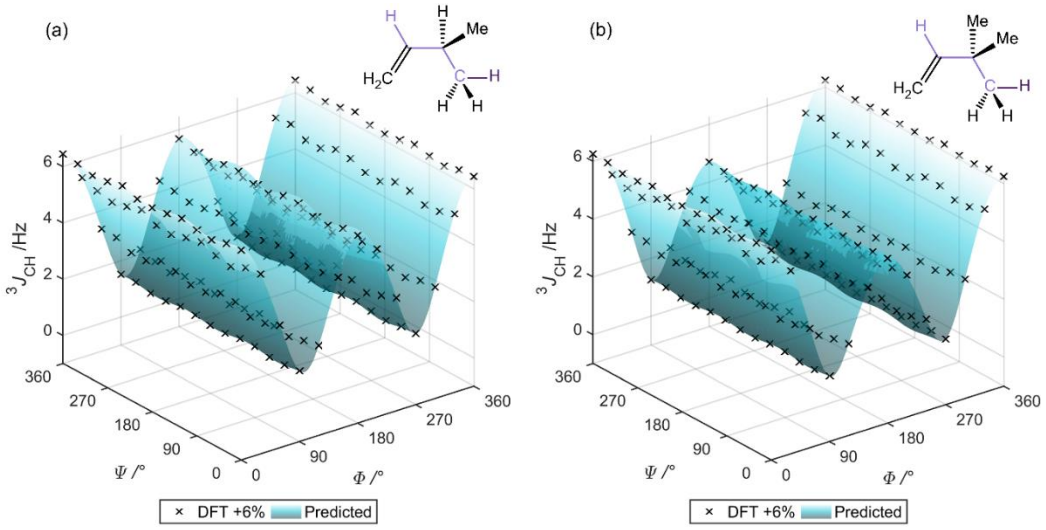


Figure IV.18 DFT-calculated $^3J_{CH}$ for (a) isopentene (VIN050) and (b) neohexene (VIN052) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

In summary, the $^3J_{CH}$ calculated for these fragments with external α, β or γ-vinyl groups show a complex relationship between the coupling constant and the dihedral angles Φ and Ψ, even in the absence of substituent effects. However, $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 64 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

IV.2.4. Internal α-β vinyl carbons, HCC=C

The effect of an internal α-β double bond on $^3J_{CH}$ was explored for 32 fragments (Figure IV.19) formed from all combinations of five proton and/or methyl substituents (VIN065-VIN096, Appendix 4). These fragments had only the Φ dihedral angle between the coupled nuclei varied, while Ψ is fixed by the presence of the internal double bond.

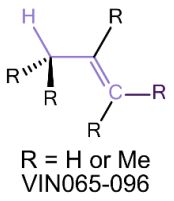


Figure IV.19 Internal α-β vinyl coupling pathway including substituent combinations, (Appendix 4).

Figure IV.20 shows the 13 $^3J_{CH}$ calculated by DFT for propene (VIN065) with Φ = 0-360° in 30° steps. The $^3J_{CH}$ show a dependence on Φ which is similar to the Karplus relationship; the maximum at 0/360° is smaller than that at 180° and the minimum $^3J_{CH}$ was at 90/270°. However, the minimum $^3J_{CH}$ do not approach 0 Hz, instead the minimum $^3J_{CH}$ was 4.5 Hz.

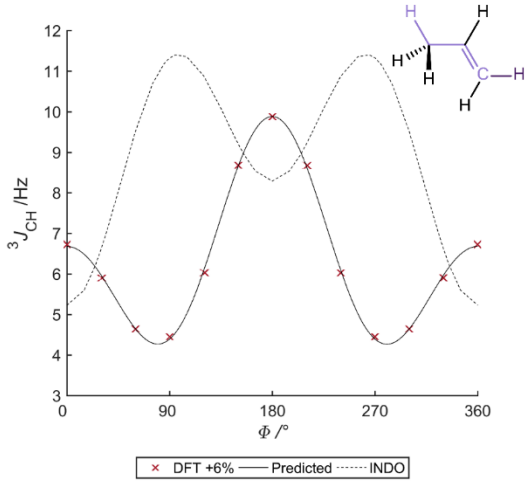


Figure IV.20 DFT-calculated $^3J_{CH}$ for propene (VIN065) dependant on the dihedral angle Φ, red 'x' data points. The solid black line indicates $^3J_{CH}$ predicted with Equation IV.5. The dashed black line indicates $^3J_{CH}$ calculated using INDO, reported by Wasylishen and Schaefer in 1973.[170]

This dependence of $^3J_{CH}$ on Φ calculated by DFT for propene differs from that found by Wasylishen and Schaefer in 1973 (Figure IV.20, dashed black line)[170] using the semi-empirical INDO method to calculate the FC contribution to $^3J_{CH}$. The INDO calculations suggested that $^3J_{CH}$ for 90/270° were at maxima rather than minima. The experimentally measured $^3J_{CH}$ of 6.7 ±0.3 Hz[171] was in good agreement with the conformationally averaged value of 6.3 Hz calculated by this work, and an improvement on the 9.0 Hz calculated by INDO, of which the poor performance is attributed to overestimation of the π-contribution to the coupling, indicated by the maximum $^3J_{CH}$ at 90/270°.[172]

The addition of an α-methyl group *trans* to the γ-carbon (Figure IV.21(a)) did not significantly affect the $^3J_{CH}$ relative to those calculated for propene; for example, the $^3J_{CH}$ ranges of 4.6-9.8 Hz for (E)-2-butene (VIN073) and 4.5-9.9 Hz for propene (VIN065) and the positions and magnitudes of the maxima/minima were unchanged. However, the addition of a *cis* α-methyl group (Figure IV.21(b)) showed a reduction in $^3J_{CH}$ for the two maxima at Φ = 180° (8.8 Hz) and Φ = 0/360° (5.1 Hz). The minimum $^3J_{CH}$ also showed a small decrease from 4.6 Hz (Φ = 90/270°, VIN073) to 4.3 Hz (Φ = 60/300°, VIN081).

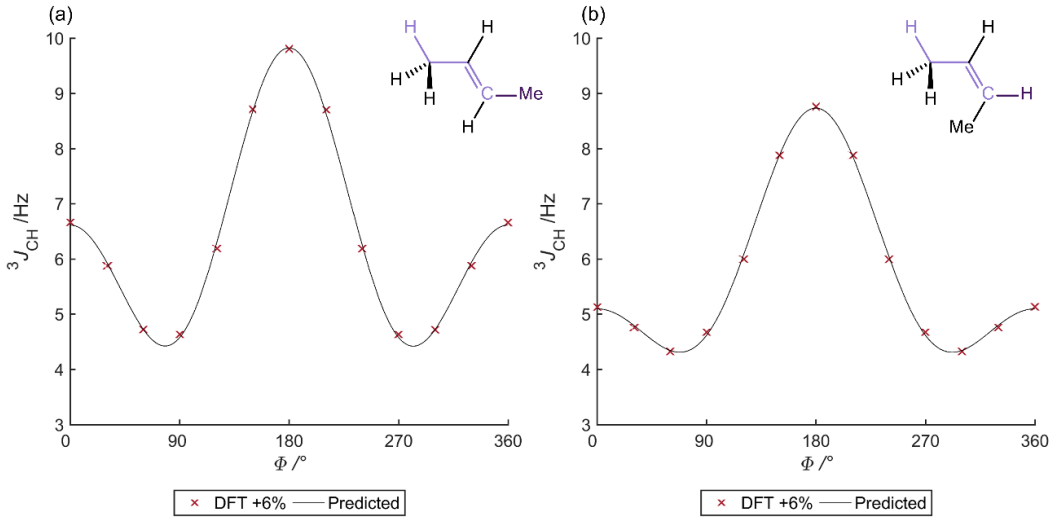


Figure IV.21 DFT-calculated $^3J_{CH}$ for (a) (E)-2-butene (VIN073) and (b) (Z)-2-butene (VIN081) dependant on the dihedral angle Φ, red 'x' data points. The black line indicates $^3J_{CH}$ predicted with Equation IV.5.

Equation IV.5, a simplification of Equation III.1, was identified for predicting $^3J_{CH}$ as a function of Φ. A linear least squares regression[168] was used to determine the C_i and U_i coefficients for the 32 fragments formed combining five proton and/or methyl substituents (VIN065-VIN096, Appendix 4).

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{3} C_i \cos(i\Phi) + \sum_{i=1}^{4} U_i \sin(i\Phi) \qquad \text{Equation IV.5}$$

The cosine terms in Equation IV.5 correspond directly to those in the Karplus equation (Equation I.13). The additional sine terms were introduced to allow an asymmetric relationship between $^3J_{CH}$ and Φ (about Φ = 180°). The capacity to fit an asymmetric relationship between $^3J_{CH}$ and Φ was necessary for fragments containing single γ-methyl groups such as (2Z)-3-methyl-2-pentene (VIN086), Figure IV.22(b). For this fragment the variation in $^3J_{CH}$ is small from 0-90° ($^3J_{CH}$ ~4 Hz), while for 270-360° there is a minimum $^3J_{CH}$ of 3.1 Hz at 300°.

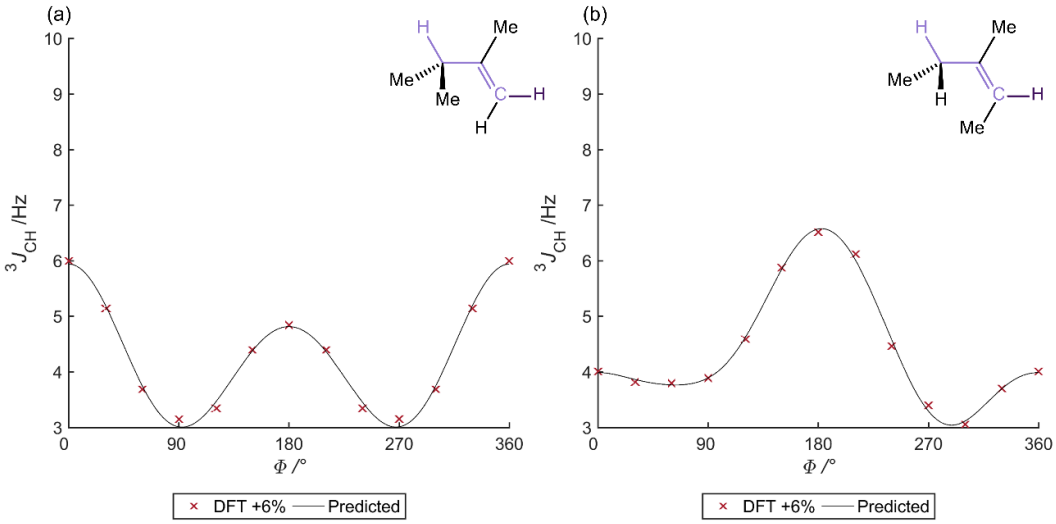


Figure IV.22 DFT-calculated $^3J_{CH}$ for (a) 2,3-dimethyl-1-butene (VIN072) and (b) (2Z)-3-methyl-2-pentene (VIN086) dependant on the dihedral angle Φ, red 'x' data points. The black line indicates $^3J_{CH}$ predicted with Equation IV.5.

Aside from the asymmetry observed for fragments with γ-methyl substituents, the substituted fragments show further deviation from Karplus-like behaviour, for example in Figure IV.22(a) the 2,3-dimethyl-1-butene (VIN072) fragment showed a maximum $^3J_{CH}$ at Φ = 0/360° rather than 180°. Additionally, for all combinations of methyl and proton substituents the minimum $^3J_{CH}$ did not approach 0 Hz. Instead 3.1 Hz was the minimum $^3J_{CH}$ observed in fragments with a single α, β and γ-methyl substituent (VIN086 and VIN084, Appendix 4). There was an average decrease of ~0.3 Hz in the total range of $^3J_{CH}$ for each methyl substituent added, however the smallest range of $^3J_{CH}$ was 3.9-6.3 Hz for (2E)-4-methyl 2-pentene (VIN079) with a *trans* α-methyl group and two γ-methyl groups.

IV.2.5. Internal β-γ vinyl carbons, HC=CC

The effect of an internal β-γ double bond on $^3J_{CH}$ was explored for 32 fragments (Figure IV.23) formed from all combinations of five proton and/or methyl substituents (VIN097-VIN128, Appendix 4). These fragments had only the Ψ dihedral angle between an α-substituent and the γ-carbon varied, while Φ is fixed by the presence of the internal double bond, forming fragments with the coupled H-C nuclei *cis* or *trans* to each other.

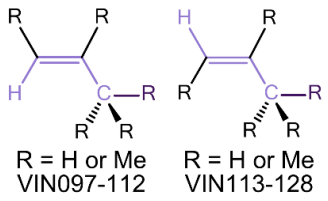


Figure IV.23 Internal β-γ vinyl coupling pathway including substituent combinations, (Appendix 4).

Figure IV.24 shows the 13 $^3J_{CH}$ calculated by DFT for propene with Ψ = 0-360° in 30° steps for the *cis* (VIN097) and *trans* (VIN113) $^3J_{CH}$. The $^3J_{CH}$ show a dependence on Ψ which is similar to that of propane (HC001, section III.2.1); with maxima at 0/360°, 120° and 240° where an α-proton substituent eclipsed with the γ-carbon and the minima at 120°, 180° and 300° where an α-proton substituent anti to the γ-carbon. The Ψ-effect for the *trans* $^3J_{CH}$ (±0.5 Hz) was larger than that observed for the *cis* $^3J_{CH}$ (±0.3 Hz). The experimentally measured couplings of 7.6 ±0.3 Hz for the *cis* $^3J_{CH}$, and 12.6 ±0.3 Hz for the *trans* $^3J_{CH}$[171] were in good agreement with the conformationally averaged values of 7.9 and 12.3 Hz calculated by this work.

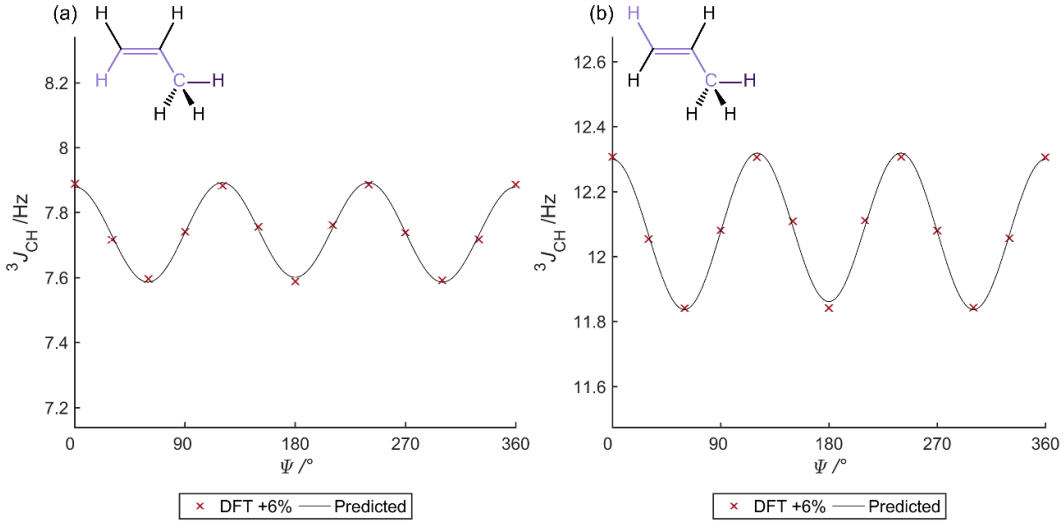


Figure IV.24 DFT-calculated $^3J_{CH}$ for propene (a) γ-proton and α-carbon *cis* (VIN097) and (b) γ-proton and α-carbon *trans* (VIN113) dependant on the dihedral angle Ψ, red 'x' data points. The black line indicates $^3J_{CH}$ predicted with Equation IV.6.

Equation III.1 was simplified to Equation IV.6 for predicting $^3J_{CH}$ as a function of Ψ. The C_j coefficients were determined from a linear least squares regression[168] and the resulting equations plotted in Figure IV.24 for the *cis* (a) and *trans* (b) $^3J_{CH}$ in propene, showing a good fit to the DFT-calculated $^3J_{CH}$ (RMSD <0.1 Hz).

$$^3J_{CH} = \sum_{j=0}^{3} C_j \cos(j\Psi) \qquad \text{Equation IV.6}$$

Equation IV.6 also performed well (RMSD ≤0.1 Hz) for the 30 β-γ vinyl fragments formed from introducing 1-5 methyl groups as substituents (VIN097-VIN128, Appendix 4). None of the fragments showed $^3J_{CH}$ with asymmetric behaviour about Ψ = 180°. Therefore, Equation IV.6 required only the cosine terms shown, and no sine terms.

The number of α-methyl substituents (Figure IV.25) effected the relationship between $^3J_{CH}$ and Ψ in an analogous manner to fragments with a saturated coupling pathway (Figure III.7, for

fixed Φ). In particular the maxima occur when the α-proton and the γ-carbon are in an eclipsed geometry, Ψ = 120/240° for a single α-methyl (Figure IV.25(a)) and Ψ = 0/360° for two α-methyl groups (Figure IV.25(b)). The isopentene fragments also show a second maximum at Ψ = 180° where the α-proton is anti to the γ-carbon.

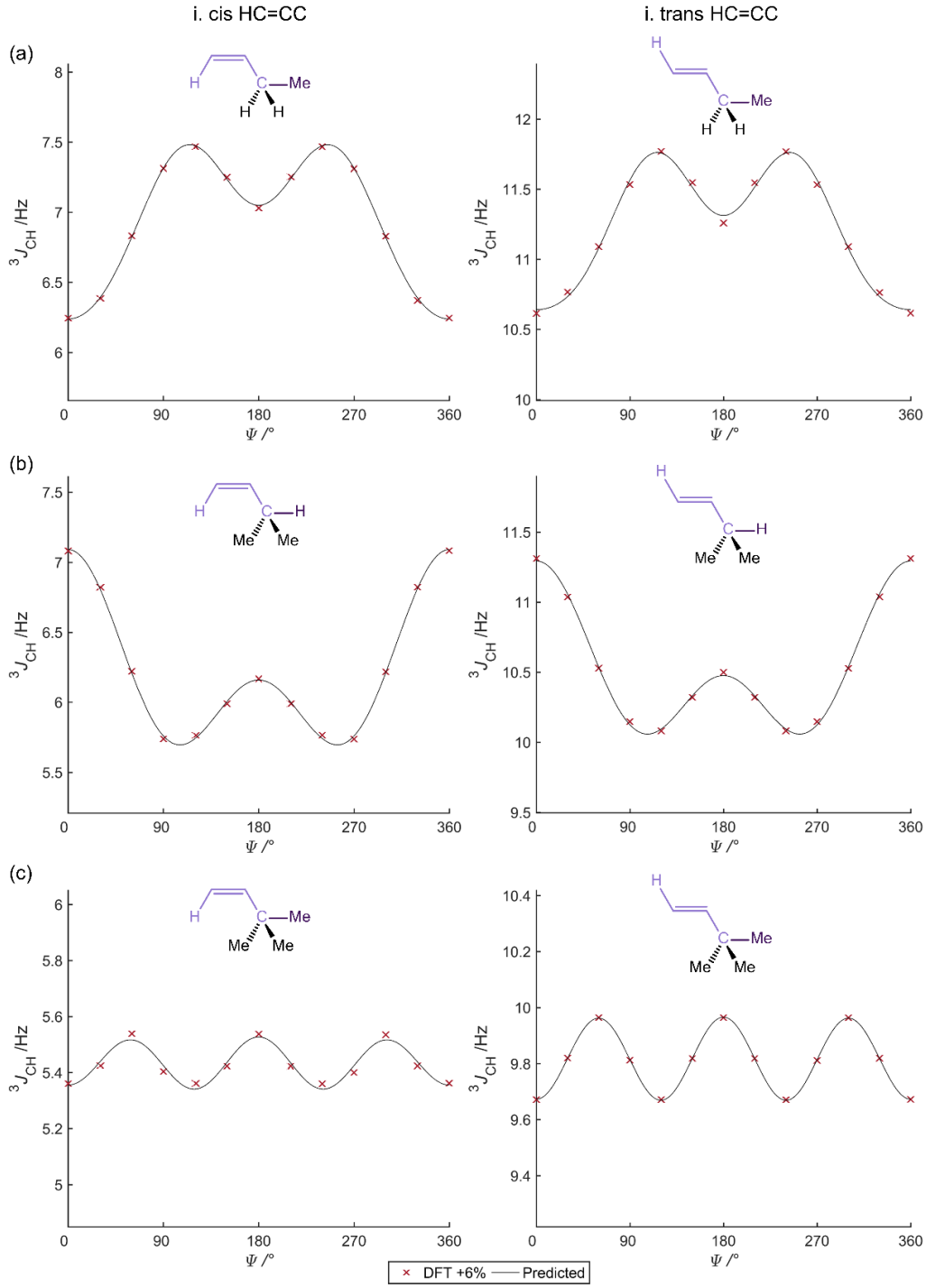


Figure IV.25 DFT-calculated $^3J_{CH}$ for (a) 1-butene, (b) isopentene and (c) neohexene for i. *cis* γ-proton and α-carbon (VIN101/VIN099/VIN103) and for ii. *trans* γ-proton and α-carbon (VIN117/VIN115/VIN119) dependant on the dihedral angle Ψ, red 'x' data points. The black line indicates $^3J_{CH}$ predicted with Equation IV.6.

The comparison of the *cis* and *trans* $^3J_{CH}$ in Figure IV.26 shows the average $^3J_{CH}$ coupling for each of the 32 combinations of 1-5 proton and methyl substituents. The $^3J_{CH}$ for propene (VIN097/VIN113) are clearly separable when considering a typical experimental error of 0.5 Hz (Table II.3) with the *trans* $^3J_{CH}$ > *cis* $^3J_{CH}$. However, for molecules such as 3,4-dimethyl 2-pentene (VIN108/VIN124) with two α, one β and one γ-methyl group the magnitude of $^3J_{CH}$ overlaps as Ψ varies. This has the potential of inverting the expected trend of *trans* $^3J_{CH}$ > *cis* $^3J_{CH}$ depending on the conformation. Therefore, the application of this $^3J_{CH}$ to structural assignment of E/Z alkenes requires careful consideration of both Ψ and the substituent pattern in addition to the accuracy of the experimental measurement.

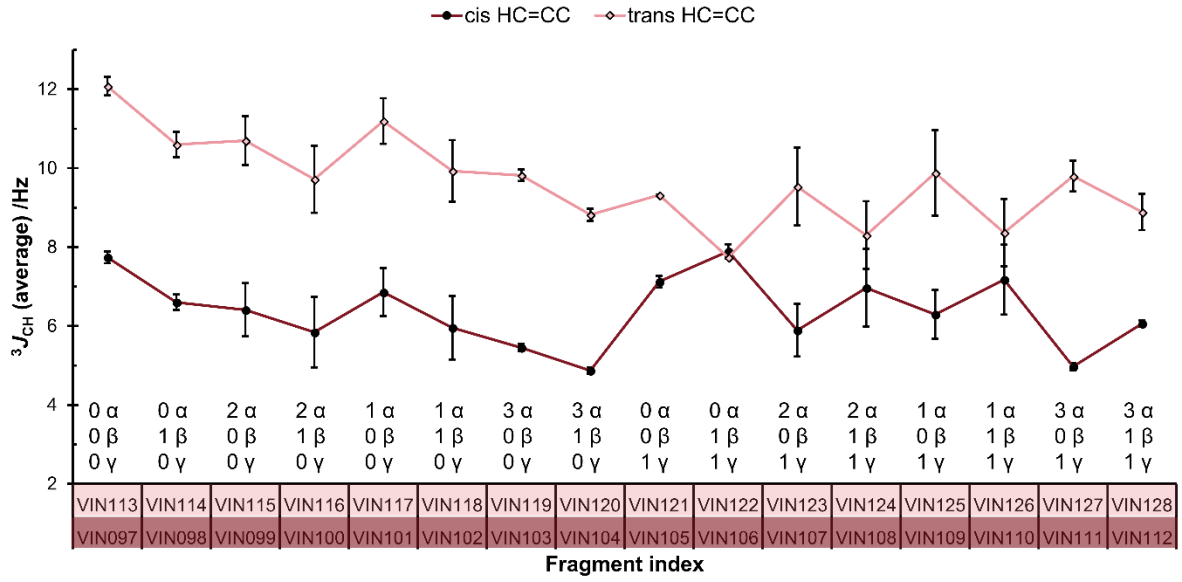


Figure IV.26 Average DFT-calculated *cis* (dark) and *trans* (light) $^3J_{CH}$ for fragments with a β-γ vinyl group and the number of α/β/γ-methyl groups indicated. The error bars indicate the variation due to Ψ.

IV.3. HCCC coupling pathway: carbonyls

The relationship between $^3J_{CH}$, coupling pathway, substituent pattern and the dihedral angles Φ and Ψ was explored for HCCC coupling pathways containing α, β or γ carbonyl carbons, Figure IV.27. $^3J_{CH}$ were calculated by DFT as a function of the dihedral angles Φ and Ψ for 64 fragments containing α, β or γ carbonyl carbons and all possible combinations of proton/methyl groups as substituents (CA001-CA064, Appendix 4)

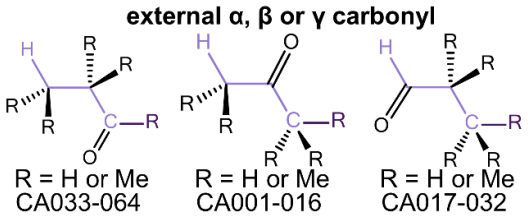


Figure IV.27 Subtypes of carbonyl containing HCCC fragments, with Fragment Index (Appendix 4).

IV.3.1. External α-carbonyl group

The relationship between $^3J_{CH}$ and the dihedral angles Φ and Ψ is shown in Figure IV.28 for propanal (CA033) a fragment with an α-carbonyl carbon and proton substituents. The relationship between $^3J_{CH}$ and Φ shows a typical Karplus-like behaviour (for fixed Ψ), with the maxima at Φ = 180° larger than those at Φ = 0/360° and symmetrical behaviour about Φ = 180°. The range of $^3J_{CH}$ (-0.1-15.4 Hz) was slightly larger than that for the 1-butene (VIN001) fragment (-0.1-12.5 Hz), which has a CH_2 group in the place of the O, as described in section IV.2.1. The $^3J_{CH}$ (Figure IV.28) show maxima at Ψ = 0/360° where the α-proton is eclipsed with the γ-carbon (for Φ = 0/360 and 180°) and at Ψ = 180° where the α-CH_2 group is eclipsed with the γ-carbon.

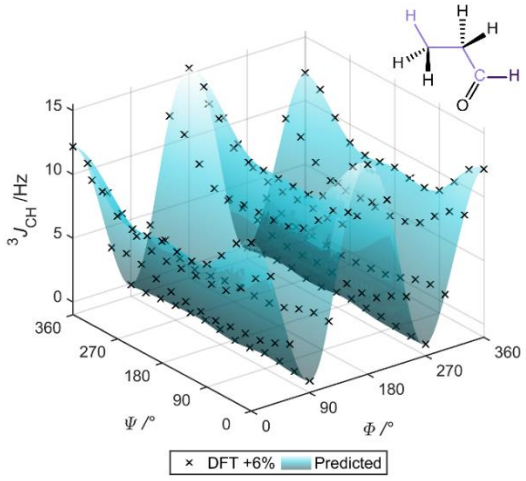


Figure IV.28 DFT-calculated $^3J_{CH}$ for propanal (CA033) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

In comparison to the external α-vinyl fragments described in section IV.2.1, the oxygen substituent results in a larger ±2.4 Hz Ψ-effect (compared to ±1.3 Hz) and a larger total range of $^3J_{CH}$. The addition of methyl substituents (CA033-CA064, Appendix 4) also led to a 1.0 Hz decrease in the total range of $^3J_{CH}$, slightly larger than the 0.9 Hz of the vinyl fragments. The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 32 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

IV.3.2. External β-carbonyl group

The simplest fragment with an external β-carbonyl coupling pathway that $^3J_{CH}$ were calculated for was acetone (CA001), shown in Figure IV.29. $^3J_{CH}$ and Φ for this molecule showed a typical Karplus-like behaviour (for fixed Ψ), with the maximum at Φ = 180° larger than that at Φ = 0/360° and symmetrical behaviour about Φ = 180°. However, the minimum $^3J_{CH}$ of -0.8 Hz (at Φ = 90°) was smaller than the 0 Hz typical of Karplus.

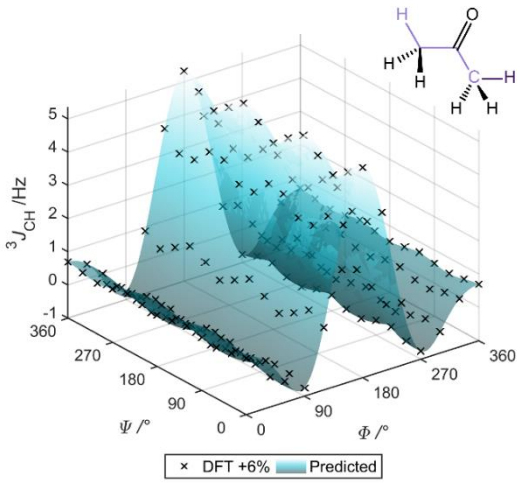


Figure IV.29 DFT-calculated $^3J_{CH}$ for acetone (CA001) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

The total $^3J_{CH}$ range (-0.8-5.33 Hz) was smaller than that for the comparable β-vinyl fragment (1.1-8.6 Hz, isobutene, VIN033) described in section IV.2.2. The introduction of oxygen to the external β double bond led to a decrease in the total $^3J_{CH}$ range, in comparison to the increase found for an α-oxygen in the previous section. The Ψ dependence of $^3J_{CH}$ in Figure IV.29 shows three maxima at Ψ = 0/360°, 120° and 240° where an α-proton is eclipsed with the γ-carbon (for fixed Φ). This effect is analogous to that observed for propane, which also contains three α-protons. The maximum Ψ-effect of ±0.4 Hz was also the same as propane (HC001).

In comparison to the external β-vinyl fragments described in section IV.2.2, the oxygen substituent resulted in a slightly larger decrease (0.4 Hz per methyl group) in the total range of $^3J_{CH}$ on addition of methyl substituents (CA001-CA016, Appendix 4). The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 16 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

IV.3.3. External γ-carbonyl group

The relationship between $^3J_{CH}$ and Φ and Ψ for the external γ-carbonyl coupling pathway is shown in Figure IV.30 for the propanal (CA017) fragment. Unlike the 1-butene (VIN049) fragment with a γ-vinyl group, the relationship between $^3J_{CH}$ and Φ shows a typical Karplus-like behaviour, whereas the 1-butene (Figure IV.16) showed maxima at Φ = 180° and 0/360° of the same magnitude.

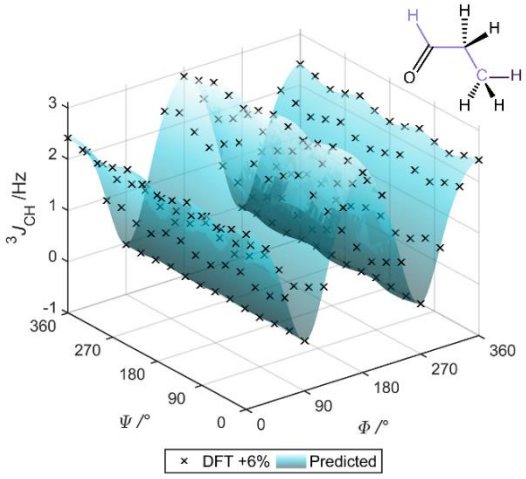


Figure IV.30 DFT-calculated $^3J_{CH}$ for propanal (CA017) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surface indicates $^3J_{CH}$ predicted with Equation III.1.

The total range of $^3J_{CH}$ for propanal (CA017) was 0.0-3.1 Hz, the smallest variation of the six vinyl and carbonyl fragments with proton-only substituents. The dependence of $^3J_{CH}$ on Ψ corresponds to the typical effect of three α-protons with maxima at Ψ = 0/360°, 120° and 240° (for fixed Φ). As was found for the γ-vinyl fragment (VIN049), the maximum Ψ-effect was ±0.2 Hz, smaller than the ±0.4 Hz observed for propane (HC001).

In comparison to the external γ-vinyl fragments described in section IV.2.3, the oxygen substituent resulted in a slightly larger increase (0.3 Hz per methyl group) in total range of $^3J_{CH}$ on addition of methyl substituents (CA017-CA032, Appendix 4). For both the γ-vinyl and γ-carbonyl the effect of adding methyl groups deviates from the α and β fragments as the total range of $^3J_{CH}$ increases rather than decreases. The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 16 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

In summary, the $^3J_{CH}$ calculated for these fragments with external α, β or γ-carbonyl groups show a complex relationship between the coupling constant and the dihedral angles Φ and Ψ. However, $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 64 fragments with proton/methyl substituents, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

IV.4. HCXC coupling pathway: β heteroatoms

The relationship between $^3J_{CH}$, coupling pathway, substituent pattern and the dihedral angles Φ and Ψ was also explored for coupling pathways containing a heteroatom in the β position of the coupling pathway (Figure IV.31). Fragments containing a β-oxygen are particularly relevant in the structural determination of carbohydrates; however, as reviewed by Coxon in 2009,[107] work to generate general empirical relationships has focussed on the effect of Φ, forming

analogues of the $^3J_{HH}$ Haasnoot-Altona equation[108] to account for the effect of substituents as a function of their electronegativities[173]. Therefore, it was of interest to explore the impact of interactions between α and γ-substituents in addition to the Ψ dihedral angle.

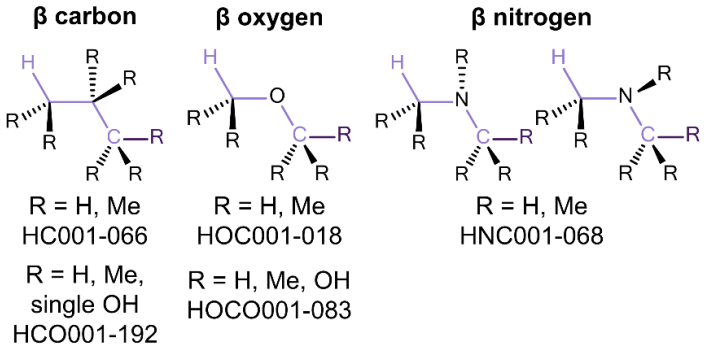


Figure IV.31 Subtypes of HCXC fragments, including Fragment Index (Appendix 4).

The fragments containing β-heteroatoms (Figure IV.32(b/c)) deviated from that of propane (Figure IV.32(a)) in several ways. Firstly, the minimum $^3J_{CH}$ were negative, -0.6 Hz for the β-nitrogen and -0.5 Hz for the β-oxygen, while for propane the minimum $^3J_{CH}$ is 0.0 Hz. The maximum $^3J_{CH}$ was also larger for the ether and amine fragments (13.6 Hz for both), resulting in a ~2 Hz larger total range of $^3J_{CH}$ than propane. Figure IV.32(c) also shows that the asymmetry of the β-nitrogen results in an asymmetric relationship between $^3J_{CH}$ and Φ (about Φ = 180°). For Φ = 0-90°, where the coupled γ-proton is on the same side of the molecule as the nitrogen lone pair, the $^3J_{CH}$ are smaller (~1 Hz) than the mirrored $^3J_{CH}$ or 360-270°. However, for Φ = 90-180°, where the coupled γ-proton is still on the same side of the molecule as the nitrogen lone pair, the $^3J_{CH}$ are larger (~1 Hz) than the mirrored 270-180°. The ether and amine fragments both showed a dependence of $^3J_{CH}$ on Ψ that corresponded to the typical effect of three α-protons – maxima at Ψ = 0/360°, 120° and 240° where an α-proton is eclipsed with the γ-carbon (for fixed Φ). The maximum Ψ-effect was ±0.5 Hz for the β-oxygen, larger than the ±0.4 Hz observed for propane (HC001) and dimethyl amine (HNC001).

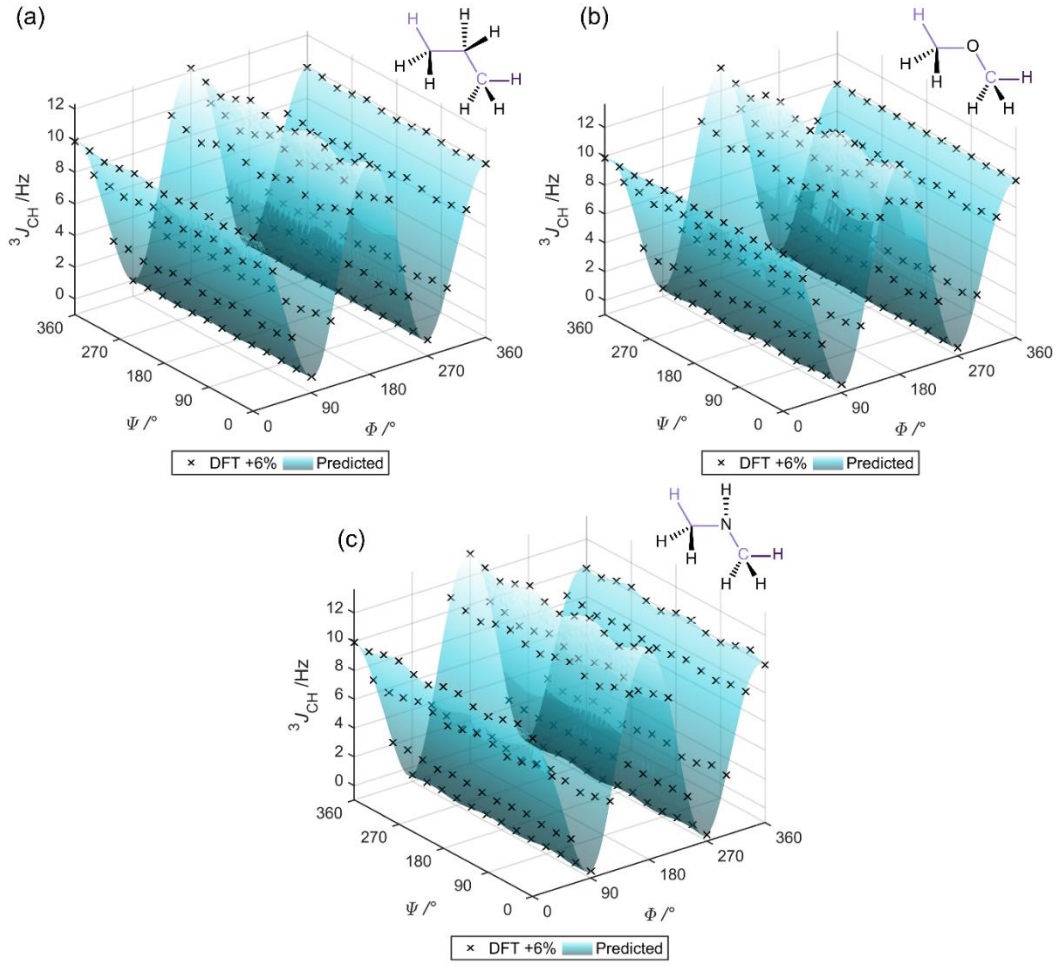


Figure IV.32 DFT-calculated $^3J_{CH}$ for (a) propane (HC001), (b) dimethyl ether (HOC001) and (c) dimethyl amine (HNC001) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1.

The effect of methyl substitution was investigated by calculating $^3J_{CH}$ by DFT as a function of the dihedral angles Φ and Ψ for the β-oxygen and β-nitrogen coupling pathways for all possible combinations of proton/methyl groups as substituents (HOC001-HOC018 and HNC001-HNC068, Appendix 4). However, the β-nitrogen proved challenging to generate all geometric combinations of Φ and Ψ as the presence of a substituent and the lone pair of the β-nitrogen led to inversion of the nitrogen centre as Φ and Ψ were varied for some fragments. Therefore, the β-nitrogen fragments will not be considered further in this thesis.

IV.4.1. HCOC coupling pathway: methyl group substituents

In Figure IV.33 the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT is related to the number of α/γ-methyl groups from the fragments. The leftmost point for each line gives the total range for a fragment containing no α-methyl groups and the rightmost point for a fragment containing three α-methyl group. The addition of an α-methyl group generally led to a decrease in the total range of $^3J_{CH}$. However, the addition of an α-methyl to fragments

containing either one γ-methyl ('●' data points) or two γ-methyl ('+' data points) led to a 1-2 Hz increase in the total $^3J_{CH}$ range (two leftmost points for each series).

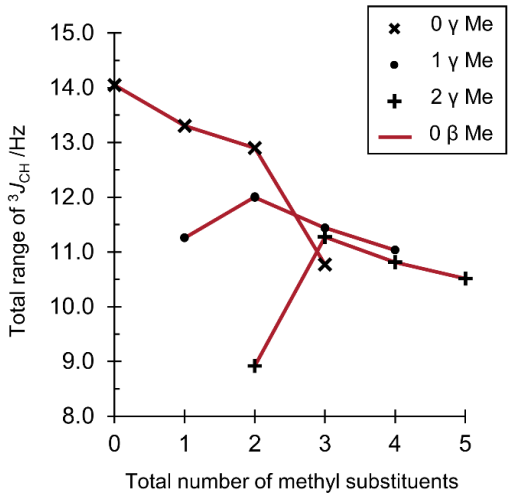


Figure IV.33 The effect of the total number of α and/or γ-methyl substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for a β-oxygen group.

Across the different fragments the minimum $^3J_{CH}$ increased from -0.5 Hz to -0.2 Hz as the number of methyl substituents increased. The relationship between $^3J_{CH}$ and the dihedral angles Φ and Ψ with changing substituent pattern were complex, however each additional methyl led to a decrease of ~0.5 Hz in the total range of $^3J_{CH}$, the same as that observed for the saturated HCCC fragments (section III.2.5). $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 18 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

IV.4.2. HCOC coupling pathway: Oxygen substituents

In section IV.1.2 the effect of oxygen substituents (OH groups) on DFT-calculated $^3J_{CH}$ was explored for the HCCC sp^3 carbon-only coupling pathway. For that coupling pathway 827 different fragments are formed from 1-7 OH substituents with 0-6 proton/methyl groups. Therefore, a smaller set of 192 fragments was targeted with all possible combinations of a single OH group and six methyl groups or protons (Appendix 4). However, for the HCOC coupling pathway the absence of β-substituents results in only 83 fragments from the combination of 1-5 OH substituents with 0-4 proton/methyl groups. The dihedral angle (μ) between the OH proton and the coupling pathway was set to 180° prior to geometry optimisation for all combinations of Φ and Ψ for each fragment.

Figure IV.34 shows the total range of $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) calculated by DFT related to the number of α/γ-OH groups from fragments containing no methyl substituents. The leftmost point for each line gives the total range for a fragment containing no α-OH groups and the rightmost point for a fragment containing three α-OH group; the addition of an α-methyl

group generally led to an increase in the total range of $^3J_{CH}$. However, the addition of three α-OH to fragments containing no γ-OH groups ('x' data points) led to a ~5 Hz decrease in the total range.

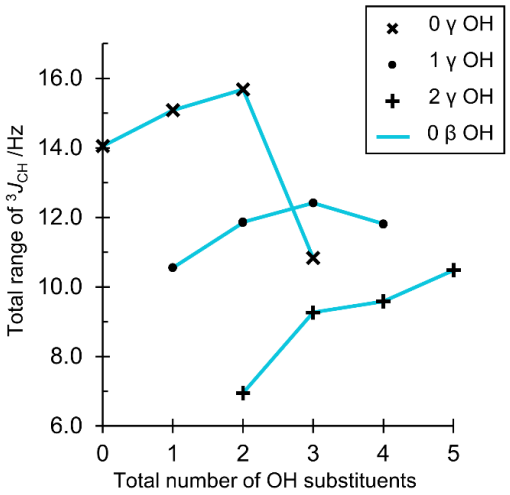


Figure IV.34 The effect of the total number of α and/or γ OH substituents on the total range of DFT-calculated $^3J_{CH}$ (maximum $^3J_{CH}$ - minimum $^3J_{CH}$) for a β-oxygen group.

Despite the typical increase in the total range of $^3J_{CH}$ (0.4-1.0 Hz per OH) for each additional α-OH group the γ-OH groups led to a ~2.5 Hz decrease, resulting in a decrease in the total range of $^3J_{CH}$ of ~0.5 Hz per OH group, similar to than that observed for methyl substituents. These fragments with mixed proton, methyl and OH substituents led to complex relationships between Φ and Ψ, such as HOCO058 (Figure IV.35) with an α-methyl group and two γ OH groups. The dependence of $^3J_{CH}$ on Φ for the fragment shows Karplus-like behaviour for Ψ = 0/360° and deviation from Karplus-like behaviour for Ψ = 180°. For the substituent combinations explored the minimum $^3J_{CH}$ remains <0 Hz (-0.4 ± 0.1 Hz), while the lowest maximum $^3J_{CH}$ is ~0.8 Hz larger than that for the HCCC coupling pathway (7.6Hz, HOCO060, three α-methyl and two γ-OH groups, Appendix 4). The $^3J_{CH}$ predicted with the 75-term variant of Equation III.1 (Table III.2(k)), fitted for each of the 83 fragments, showed good agreement with the DFT-calculated $^3J_{CH}$ (RMSD ≤0.1 Hz, Appendix 4).

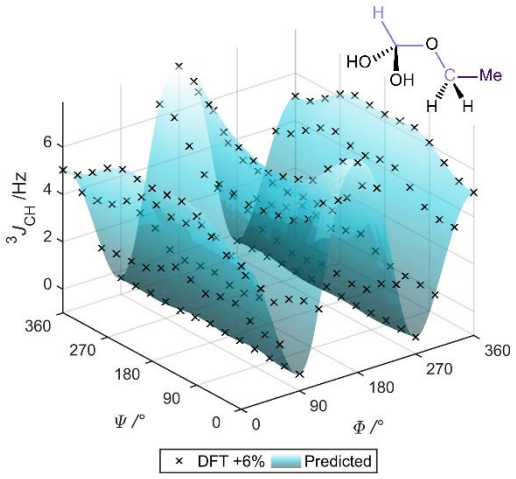


Figure IV.35 DFT-calculated $^3J_{CH}$ for ethoxymethandiol (HOC058) dependant on the dihedral angles Φ and Ψ, black 'x' data points. The transparent surfaces indicate $^3J_{CH}$ predicted with Equation III.1.

IV.5. Summary

Chapter III and Chapter IV focussed on the relationship between DFT-calculated $^3J_{CH}$ and the dihedral angles Φ and Ψ. This data was used to produce a library of 556 sets of coefficients for empirically predicting $^3J_{CH}$ with Equation III.1 (Φ and Ψ dependant), Equation IV.5 (Φ dependant) or Equation IV.6 (Ψ dependant) that are distinguished by the coupling pathway and substituent pattern of the fragments used to generate them.

The fragments (Figure IV.1) consisted of twelve different coupling pathways including saturated/unsaturated centres and β-heteroatoms; they also covered varied α, β and γ-substituents, with the primary focus on methyl groups. The mixed α, β and γ-substituents were found to have a substantial impact on $^3J_{CH}$, giving a large range in the maximum $^3J_{CH}$ calculated. Additionally, the complex symmetry of the molecules formed from multiple substituents typically led to an asymmetric relationship between $^3J_{CH}$ and Φ and/or Ψ about 180° and further deviation from typical Karplus-like behaviour in terms of the exact positions (Φ and Ψ values) and relative magnitudes of the maxima and minima.

However, by careful consideration of the terms included, the equations developed predicted the behaviour found for the DFT-calculated $^3J_{CH}$ to a reasonable degree of accuracy (RMSD ≤0.1 Hz) for the structural variations investigated. The validation of these equations by comparison to experimentally measured $^3J_{CH}$ and calculated $^3J_{CH}$ is described in Chapter V and Chapter VI respectively.

Chapter V. Validation of empirical $^{3}J_{CH}$ prediction against experimentally measured $^{3}J_{CH}$

V.1. Introduction

In Chapter III and Chapter IV, empirical relationships between three-bond proton-carbon scalar coupling constants ($^{3}J_{CH}$) and the dihedral angles Φ and/or Ψ were developed for twelve coupling pathways (Figure IV.1) in combination with different substituent patterns, resulting in a library of 556 sets of coefficients (Appendix 4) for Equation III.1, Equation IV.5 and Equation IV.6. These equations predict the DFT-calculated $^{3}J_{CH}$ for whole molecules to a reasonable degree of accuracy (RMSD ≤0.1 Hz) for the structural variations investigated. However, the efficacy of the equations in practice depends on the prediction of experimental data.

Therefore, long range proton-carbon scalar coupling constants ($^{n>1}J_{CH}$) were measured experimentally using the methods described in Chapter II for compounds (Figure V.1) that were rigid, or contained rigid portions, such that the experimental $^{3}J_{CH}$ could be correlated directly with structural parameters obtained from DFT geometry optimisation. The accuracy of $^{n>1}J_{CH}$ calculated by DFT was also assessed (section V.2), providing a comparison for the accuracy achieved with empirical equations (section V.3) for the same set of molecules. Where the experimentally measured data failed to cover the breadth of equations, available literature $^{n>1}J_{CH}$ were also used to validate their performance (section V.5).

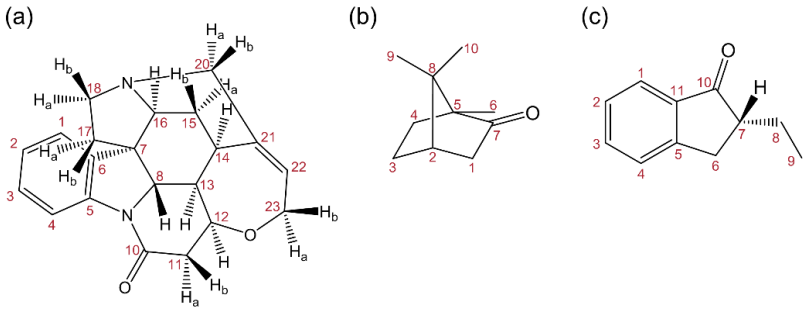


Figure V.1 Labelling system used for (a) strychnine, (b) camphor and (c) 2-ethyl-1-indanone. The geminal protons in camphor and 2-ethyl-1-indanone (defined with the *R*-enantiomer shown) are distinguished using Ha (*pro*-*R*) and Hb (*pro*-*S*), see Appendix 1 for full ^{1}H and ^{13}C assignment.

V.2. Experimental measurement of $^{n>1}J_{CH}$

The $^{n>1}J_{CH}$ were measured for strychnine, camphor and 2-ethyl-1-indanone (Figure V.1) using three orthogonal methods: coupled ^{13}C, EXSIDE[151] and IPAP-HSQMBC[154,157]. As discussed in Chapter II, each method has different advantages relative to the others; EXSIDE, for example, offers the simplest extraction method (from a scaled doublet in F1), but it is limited to ^{1}H that can be selected separately from mutually coupled ^{1}H. The broadband technique IPAP-HSQMBC is not limited by selectivity requirements, however overlap between ^{1}H that couple

to either the same ^{13}C (or ^{13}C that are close in chemical shift) complicates the extraction of $^{n>1}J_{CH}$. Coupled ^{13}C spectra require the most complex analysis to extract the $^{n}J_{CH}$, however they are not limited by overlap in the ^{1}H spectrum. The combination of these techniques reduced the impact of their individual weaknesses and maximised the number of measurable couplings (Table V.1(a-c)), leading to an average 40% increase in the number of measured couplings compared to using any single technique.

Table V.1 Comparison of the number of experimentally measured $^{n>1}J_{CH}$ for (a) strychnine, (b) camphor and (c) 2-ethyl-1-indanone.

	(a) Strychnine $^{n>1}J_{CH}$	(b) Camphor $^{n>1}J_{CH}$	(c) 2-ethyl-1-indanone $^{n>1}J_{CH}$
Coupled ^{13}C	44	33	47
IPAP-HSQMBC[a]	74	39	27
EXSIDE (tilted)	46	27	33
Total	99	53	53

[a] Accordion IPAP-HSQMBC for strychnine and camphor.

V.2.1. DFT-calculated $^{n>1}J_{CH}$

The calculation of $^{n>1}J_{CH}$ was performed using Gaussian 09[78] as described in Chapter II for strychnine and camphor and compared to the experimental data in order to set an approximate target for the potential accuracy of the empirical equations (which are derived from the same DFT method). The minor (2.5%) conformer of strychnine, identified by NOE measurements[174], that arises from flexibility of the 7-membered ring was not included for scalar coupling calculation due to the negligible contribution of lowly populated conformations. For the conformer ratio of 0.975:0.025 the maximum potential error in a given DFT-calculated $^{n>1}J_{CH}$ arising from not considering the minor conformer is 0.3 Hz, assuming a typical $^{n>1}J_{CH}$ range of 0-10 Hz.

A conformational search was performed with MacroModel[175] for the more flexible 2-ethyl-1-indanone using a MCMM (Monte Carlo Multiple Minimum) method[176] resulting in three conformers. DFT calculations were performed with Gaussian 09 starting with geometry optimisation and frequency calculation using mPW1PW91/6-311G(d,p); Cartesian coordinates are provided in Appendix 3. NMR calculations were performed with mPW1PW91/6-311G(d,p) using the GIAO method, including total scalar coupling constants (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms) that were corrected by 6% as described in Chapter II.

The three rotamers of the ethyl group in 2-ethyl-1-indanone, combined with two conformations of the five-membered ring (C7 above and below the plane of the ring) suggests six possible conformers. However, only three conformers were found by the molecular mechanics conformational search. These correspond to ethyl rotations with the ethyl group in a pseudo-

equatorial position and a C5-C11-C10-C7 dihedral angle of approximately -10° for the three conformers (Cartesian coordinates are provided in Appendix 3). Modifying the conformational search by increasing the energy window that conformations are retained within from 50 kJ/mol to 150 kJ/mol and the atom deviation used to eliminate redundant conformations from 0.5Å to 0.05Å found only the same initial three conformers. Manually generating the other three suggested conformers with the ethyl group pseudo-axial by setting the dihedral angle C5-C11-C10-C7 to +10° and performing a geometry optimisation with DFT (mPW1PW91/6-311G(d,p)) found the same initial three conformers (C5-C11-C10-C7 ~ -10°). This suggests that the other three possible conformations with the ethyl group in a psudo-axial position are not energy minima. Therefore, the scalar couplings across the three conformers found were weighted according to the Boltzmann distribution (Equation V.1) using the Gibbs free energies determined for each conformer by the frequency calculations; populations are provided in Appendix 3. The full computational details for strychnine, camphor and 2-ethyl-1-indanone are provided in Chapter VIII (Experimental).

$$J_{\mathrm{xy}} = \frac{\sum_i J_{\mathrm{xy}}(i)\, e^{-G_i/RT}}{\sum_i e^{-G_i/RT}} \qquad \text{Equation V.1}$$

Where, J_{xy} is the Boltzmann-averaged scalar coupling constant between nuclei x and y; $J_{xy}(i)$, the scalar coupling constant between nuclei x and y for conformer *i*; G_i, the relative Gibbs free energy of conformer *i* and T, the temperature (298 K).

Figure V.2 compares the absolute values of the DFT-calculated $^{n>1}J_{CH}$ and the experimentally measured $^{n>1}J_{CH}$ values for each molecule. The sign cannot generally be determined by the experimental methods used and so is ignored here. In Figure V.2(a-c) separating the DFT-calculated $^{n>1}J_{CH}$ and experimentally measured $^{n>1}J_{CH}$ by molecule confirms that, for this simple example, accounting for the conformational flexibility of 2-ethyl-1-indanone achieves the same accuracy in $^{n>1}J_{CH}$ calculation as the single conformers of strychnine and camphor (MAD/SD 0.4/0.5 Hz, Table V.2(a-c)). The MAD/SD for these data also agree with the 0.4/0.5 Hz (Table II.3) determined in Chapter II from the comparison of DFT-calculated $^{n>1}J_{CH}$ to the data measured from coupled ^{13}C spectra for strychnine and camphor.

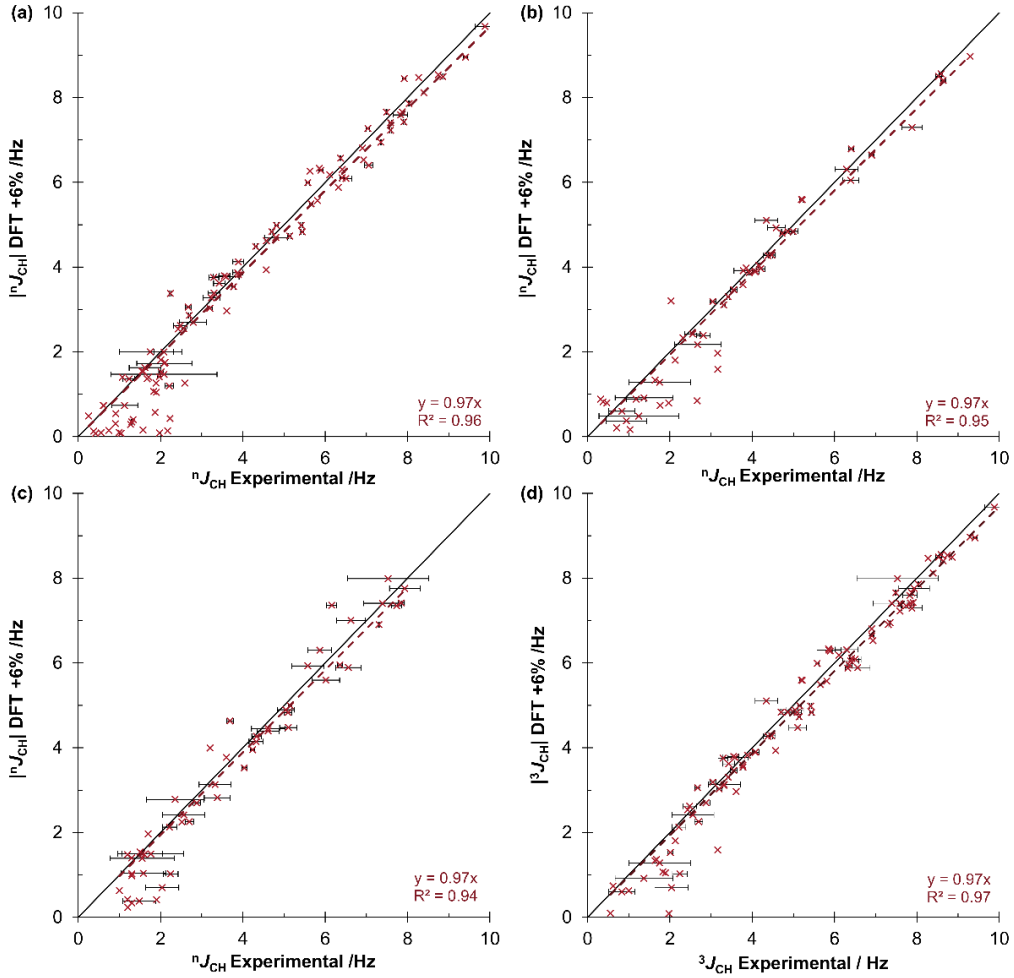


Figure V.2 Comparison of DFT-calculated $^{n>1}J_{CH}$ (mPW1PW91/6-311G(d,p), IEFPCM(chloroform)) and average experimentally measured $^{n>1}J_{CH}$ for (a) strychnine, (b) camphor and (c) 2-ethyl-1-indanone. (d) shows the combined $^{3}J_{CH}$ for all three molecules. The error bars indicate the standard deviation between the experimental methods: coupled ^{13}C, EXSIDE and accordion IPAP-HSQMBC[125] (strychnine/camphor) or IPAP-HSQMBC[154,157] (2-ethyl-1-indanone).

From the 205 experimentally measured $^{n>1}J_{CH}$ there were 102 $^{3}J_{CH}$ (Figure V.2(d)), on comparison to the DFT-calculated $^{3}J_{CH}$ the MAD/SD (0.4/0.4 Hz, Table V.2(d)) showed similar values to the $^{n>1}J_{CH}$ datasets. This data establishes a target for empirically predicted $^{3}J_{CH}$ using the same geometric parameters.

Table V.2 Comparison of DFT-calculated and average experimentally measured $^{n}J_{CH}$ for strychnine (a), camphor (b) and 2-ethyl-1-indanone (c), (d) shows the combined $^{3}J_{CH}$ for all three molecules.

	(a) Strychnine $^{n>1}J_{CH}$	**(b) Camphor** $^{n>1}J_{CH}$	**(c) 2-ethyl-1-indanone** $^{n>1}J_{CH}$	**(d) All** $^{3}J_{CH}$
Mean absolute deviation (MAD) / Hz	0.43	0.41	0.43	0.37
Standard deviation (SD) / Hz	0.53	0.53	0.52	0.44
Total number of experimental values	99	53	53	102

V.3. Empirical prediction of $^3J_{CH}$

The structural information required to empirically predict the scalar coupling constants for a molecule depends on the complexity of the equation. The equation developed by Aydin *et al.* [123], for example, is an experimental parameterisation of the Karplus equation for $^3J_{CH}$ (Equation V.2) that only requires the dihedral angle (Φ) between the coupled nuclei as an input.

$$^3J_{CH} = 4.50 - 0.87\cos\Phi + 4.03\cos 2\Phi \qquad \text{Equation V.2}$$

However, the more complex equation developed by Palermo *et al.* (Equation V.3)[10] is a function both Φ and the β/γ-substituent electronegativities and orientations. Therefore, using Equation V.3 to predict $^3J_{CH}$ for a molecule requires interpretation of the molecular structure beyond the geometric considerations (determining Φ from the molecular coordinates).

$$\begin{aligned} ^3J_{CH} = {} & P_1\cos^2\Phi + P_2\cos\Phi + P_3 \\ & + \sum \Delta\chi_{H\pm}[P_4 + P_5\cos^2(\xi_i\Phi + P_6|\Delta\chi_{H\pm}|)] \\ & + \sum \Delta\chi_{C\pm}[P'_4 + P'_5\cos^2(\xi_i\Phi + P'_6|\Delta\chi_{C\pm}|)] \end{aligned} \qquad \text{Equation V.3}$$

Equation I.20, repeated for clarity. Empirical equation from Palermo *et al.*[10] accounting for β and γ-substituent effects for $^3J_{CH}$. $P_{1\text{-}6}$ and $P'_{4\text{-}6}$ are constants. $\Delta\chi$ is the electronegativity of the substituent relative to hydrogen where H indicates γ-substituents and C β-substituents. ξi is 1 or -1 depending on the orientation of the substituent, as defined in Figure I.16.

The fragment-based equations described in this work (Chapter III, Equation III.1 and Chapter IV, Equation IV.5 and Equation IV.6) use relatively simple geometric input parameters, specifically the two dihedral angles Φ and/or Ψ. However, accounting for coupling pathway and substituent effects requires interpreting the molecular structure and hence the fragment to identify a set of coefficients from the library of >500 sets. In Figure V.3 the H1a-C3 scalar coupling in camphor is used as an example. The coupling pathway C1-C2-C3 consists of three carbons connected by single bonds, the substituents (proton/carbon) are also all connected by single bonds suggesting a fragment from the "HC001-HC066" sets of coefficients. The stereochemistry and type of substituents (proton/carbon) define the fragment as HC048 (and therefore the coefficients for Equation III.1). Fragment HC048 uses methyl groups as proxies for the α (C4, 2°), β (C8, 4°) and γ (C7, 4° carbonyl) carbon substituents. Inputting the Φ (-170°) and Ψ (-70°) values from the DFT geometry optimised structure to Equation III.1 gives a predicted $^3J_{CH}$ of 8.8 Hz which is in good agreement with the 8.6 ±0.1 Hz measured by experiment (Appendix 1).

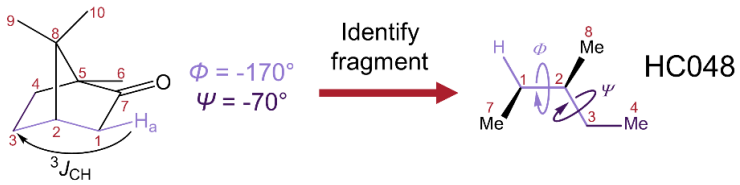


Figure V.3 Fragment based approach to predicting $^3J_{CH}$ using H1a-C3 in camphor as an example.

The workflow outlined in Figure V.4 was developed to identify suitable fragments and predict $^3J_{CH}$ from the Cartesian coordinates of a molecule. The first step was determination of the connectivity (which atoms are bonded) and an estimated bond order using GaussView5[177]. All dihedral angles between connected atoms were calculated from the Cartesian coordinates. The connectivity established which proton/carbon nuclei were separated by three bonds, the nature of the αβγ coupling pathway, and α, β and γ-substituents while also excluding proton/carbon pairs with multiple coupling pathways of three bonds or fewer. The combination of substituent type, bond order and dihedral angles were used to determine the orientation of substituents relative to the coupling pathway and therefore identify the matching fragment. Following identification of the fragment, the relevant equation and coefficients were used to predict $^3J_{CH}$ using the dihedrals *Φ* and *Ψ*.

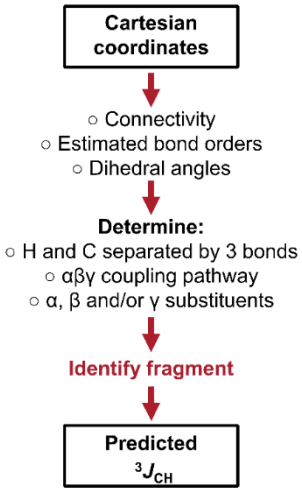


Figure V.4 Workflow for predicting $^3J_{CH}$ from the Cartesian coordinates of a molecule.

V.3.1. Comparison to experimentally measured $^3J_{CH}$

The equations developed by Aydin *et al*. (Equation V.2) and Palermo *et al*. (Equation V.3) were used to predicte $^3J_{CH}$ for strychnine, camphor and 2-ethyl-1-indanone and compared against the performance of the fragment-based approach described in this work (Equation III.1, Equation IV.5 and Equation IV.6). The workflow outlined in Figure V.4 also generates the required input parameters (*Φ*, and the substituent orientation and nature) to predict $^3J_{CH}$ by both literature equations. Figure V.5 shows the comparison of the empirically predicted $^3J_{CH}$ to the experimentally measured $^3J_{CH}$. The multiple conformers of 2-ethyl-1-indanone were

accounted for by empirical prediction of $^3J_{CH}$ from the geometric parameters (Φ and/or Ψ) of each conformer and averaging the three values using the DFT-calculated populations (Appendix 3) that were also used for the DFT-calculated $^3J_{CH}$ (Equation V.1).

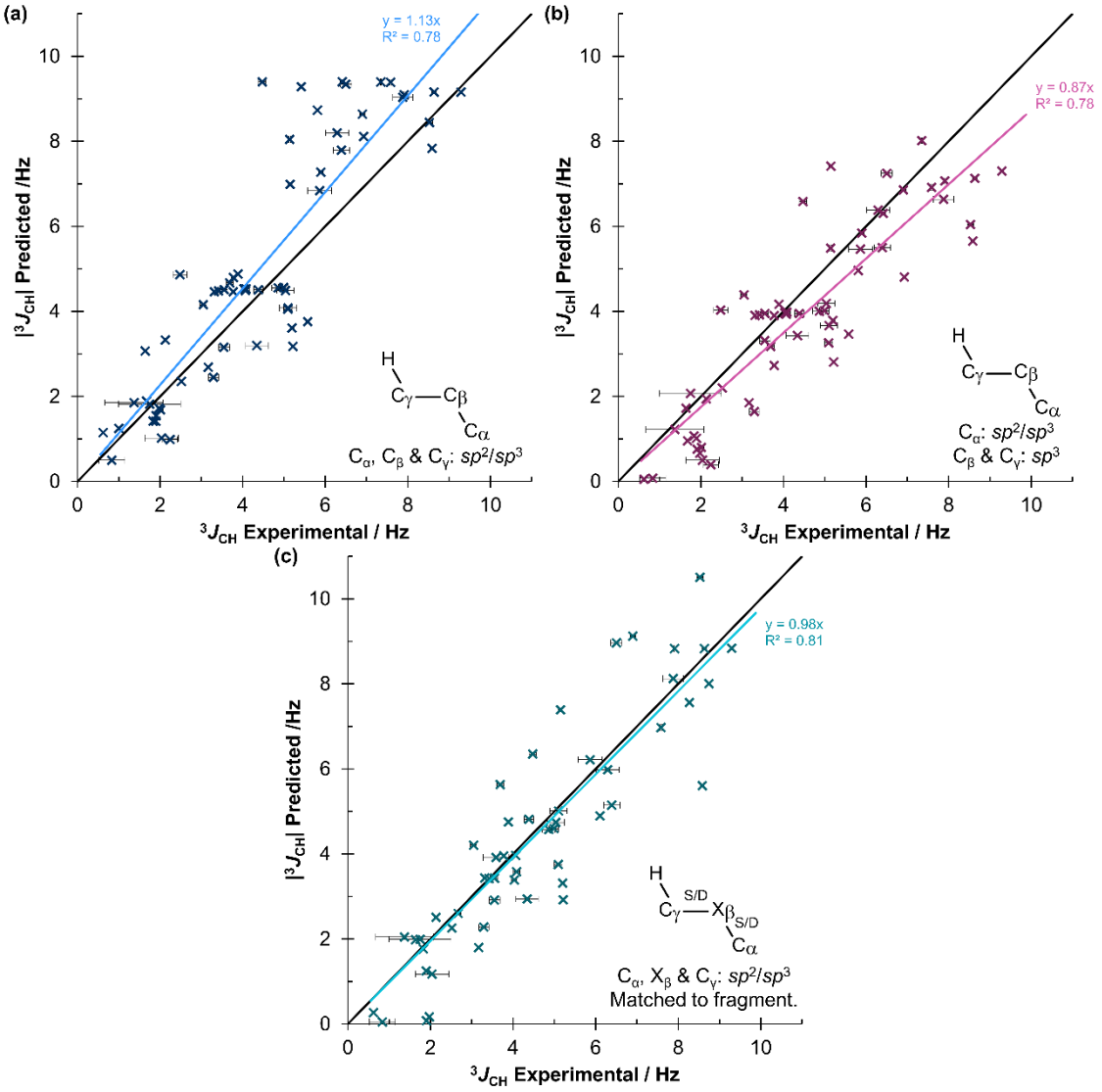


Figure V.5 Comparison of experimentally measured $^3J_{CH}$ and empirically predicted $^3J_{CH}$ by (a) Equation V.2, Aydin *et al.*[123]; (b) Equation V.3, Palermo *et al.*[10]and (c) the fragment-based approach described in this work. The solid black line in (a-c) indicates 'y = x'.

The comparison of the $^3J_{CH}$ predicted with Equation V.2 (Aydin *et al.*) to the experimentally measured values (Figure V.5(a)) shows that the predicted $^3J_{CH}$ were generally larger than the experimentally measured $^3J_{CH}$. As indicated in Figure V.5(a), Equation V.2 was applied to $^3J_{CH}$ with a coupling pathway consisting of three carbons connected by single bonds, therefore excluding $^3J_{CH}$ within the aromatic rings and $^3J_{CH}$ with β-heteroatoms that show a different dependence on Φ[123]. Figure V.5(a) shows the limitations of using a single structural parameter (Φ) to predict $^3J_{CH}$, there are groups of couplings clustered around the lines 'y = 4.5' and 'y =

9.5' where the similar values of (±)Φ predict approximately the same $^3J_{CH}$ value despite the corresponding experimental values showing 2-5 Hz variation.

The equation developed by Palermo *et al*. (Equation V.3)[10] accounts for the Φ-dependence of $^3J_{CH}$ and the effects of β/γ-substituents. The application of Equation V.3 in Figure V.5(b) was matched to that of Palermo *et al*., where the γ and β-carbons were necessarily sp^3 hybridised, but the hybridisation-state of the α-carbon was not limited. The linear trendline in Figure V.5(b) shows that the predicted $^3J_{CH}$ are generally lower than the experimentally measured $^3J_{CH,}$ (y = 0.87x). Similar to Figure V.5(a) couplings are clustered around the line 'y = 4' where similar (±)Φ values predict approximately the same $^3J_{CH}$, despite accounting for β/γ-substituent effects in addition to Φ. The experimental variation might be due to Ψ or α-substituent effects on $^3J_{CH}$, however other geometric parameters such as bond angle or bond length may also contribute.

Figure V.5(c) shows the empirical prediction of $^3J_{CH}$ by the fragment-based approach, this method calculated $^3J_{CH}$ only when one of the >500 fragments exactly matched with the coupling pathway and substituent pattern (1-bond from the coupling pathway). The groups of predicted $^3J_{CH}$ with the same value (4.0-4.5 Hz and ~9.0 Hz) despite the experimental $^3J_{CH}$ showing 2-5 Hz variation were not observed for this method. The linear trendline (y = 0.98x) approaches 'y = x', showing that there is not a systematic overestimation or underestimation of the $^3J_{CH}$.

V.3.2. Performance of DFT and empirical methods

The performance of DFT-calculated and empirically predicted $^3J_{CH}$ was evaluated by calculating the mean absolute deviation and standard deviation (MAD/SD, Table V.3) and visualising the results; Figure V.6 shows the percentage population of the deviation between the experimental and estimated $^3J_{CH}$ plotted for 0.5 Hz ranges centred on 0 Hz. The DFT-calculated $^3J_{CH}$ (Figure V.6(a)) showed the highest population of the low error central category (-0.25, 0.25 Hz] at 45% compared to 11% for Aydin *et al*., 20% for Palermo *et al*. and 21% for the fragment-based approach. DFT-calculated $^3J_{CH}$ also showed the smallest range of deviations -0.9-1.9 Hz and the lowest total population of high errors (9%, |Deviation|>0.75 Hz). This good performance is captured by the MAD/SD 0.4/0.4 Hz (Table V.3(a)), the smallest of any of the methods.

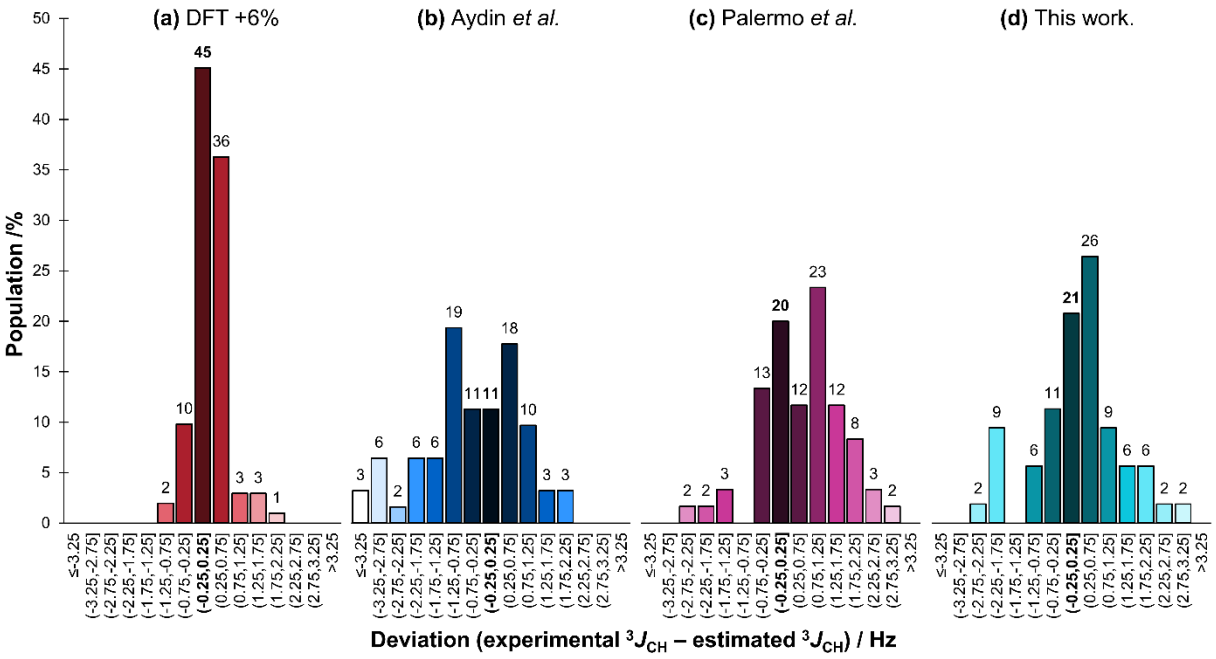


Figure V.6 The percentage population of the deviation (0.5 Hz categories) between estimated and experimentally measured $^3J_{CH}$ for all three molecules (a) DFT-calculated (b) Aydin *et al*. [123] (c) Palermo *et al*.[10] (d) this work.

In Figure V.6(b), Aydin *et al*., the most populated category (19%) was (-0.75, -1.25 Hz] and when combined with the positive category, (0.75, 1.25 Hz], this sums to 39% of the predicted $^3J_{CH}$. The method also showed the largest deviations of -4.9 Hz (C5-H2, camphor), 3.9 Hz (C14-H20b, strychnine) and 3.0 Hz (C14-H16, strychnine). The highly substituted nature of the three scalar couplings (proton, carbon and/or nitrogen substituents) clearly contributes to the deviation from the simple Φ-dependence described by Equation V.2. This method showed the highest MAD/SD (1.1/1.4 Hz, Table V.3(b)), highlighting the limitations of not directly accounting for substituents and suggesting that the equation has a limited use for 3D structural determination. Fitting Equation V.2 using $^3J_{CH}$ data from a broader range of compounds such as those with oxygen and nitrogen substituents might improve the prediction of $^3J_{CH}$ (reducing the MAD) for these compounds (strychnine, camphor and 2-ethyl-1-indanone), and this is explored in Chapter VI.

The introduction of β/γ-substituent effects to Equation V.3 by Palermo *et al*. (Figure V.6(c)) showed an increase in the population of the low error, central category (-0.25, 0.25 Hz] to 20%, however the most populated category was the (0.75, 1.25 Hz] category at 23%. The range of deviations was -2.3-2.9 Hz, a significant improvement on the maximum errors found for $^3J_{CH}$ predicted when only considering Φ. However, in comparison to the DFT-calculated $^3J_{CH}$ the higher MAD/SD (1.0/1.1 Hz, Table V.3(c)) showed a lower accuracy.

The fragment-based approach (Figure V.6(d)) includes the effects of Φ, Ψ and α, β, and γ-substituent effects, but only achieves a population of 21% in the central category, lower than the 45% of DFT (Figure V.6(a)) and comparable to the 20% of Palermo *et al.* (Figure V.6(c)). However, when considering $^3J_{CH}$ with low errors (|Deviation| ≤ 0.75 Hz), the total population increases when more effects are considered empirically: Aydin *et al*. 40%, Palermo *et al*. 45% and the fragment-based approach 58%, although the DFT-calculated $^3J_{CH}$ show significantly greater accuracy at 91%.

The final row of Table V.3 shows the total number of $^3J_{CH}$ estimated by each method for the three compounds. DFT calculates a scalar coupling value for every pair of nuclei in a molecule, therefore 129 represents the total number of proton and carbon nuclei separated by three bonds (excluding $^3J_{CH}$ with multiple coupling pathways of three bonds or less). The less restricted application of Equation V.2 (Table V.3(b)) leads to prediction of the largest number (65%) of couplings of the empirical methods. Equation V.3 (Table V.3(c)) shows a small decrease in the number of couplings that could be predicted (60%), while the fragment-based approach covered less than half (47%), highlighting the need to extend the variety of coupling pathways and substituent patterns of the fragments. When limited to scalar couplings that also have an experimental value for comparison (penultimate row of Table V.3) the differences between the empirical methods reduce (61-51%).

A more direct comparison between the empirical methods was performed by limiting the dataset to scalar couplings within the scope of all three empirical methods, with an HCCC sp^3-only carbons coupling pathway and proton/carbon substituents. These $^3J_{CH}$ were compared to the 33 scalar couplings from the subset that were measured experimentally, these numbers are included in brackets in Table V.3.

Table V.3 Comparison of estimated and experimentally measured $^3J_{CH}$ for all three molecules (a) DFT-calculated (b) Aydin *et al*.[123] (c) Palermo *et al*.[10] (d) this work. The numbers in brackets were calculated by limiting to an HCCC coupling pathway with sp^3-only carbons and proton/carbon substituents.

	(a) DFT	**(b) Aydin *et al*.**	**(c) Palermo *et al*.**	**(d) This work**
Mean absolute deviation (MAD) / Hz	0.37 (0.33)	1.14 (1.03)	0.95 (0.99)	0.88 (0.84)
Standard deviation (SD) / Hz	0.44 (0.42)	1.38 (1.32)	1.06 (1.17)	1.15 (1.15)
Number of estimated and experimental $^3J_{CH}$	102 (33)	62 (33)	60 (33)	53 (33)
Total number of estimated $^3J_{CH}$	129 (37)	84 (37)	78 (37)	60 (37)

Limiting the data to simple hydrocarbon fragments led to slight improvements in the MAD (<0.1 Hz) for all methods except that of Palermo *et al*. where the generalisation of the Equation V.3 is designed to account for the effects of a range of electronegative substituents. DFT remained the most accurate method (MAD/SD 0.3/0.4 Hz), however an additional consideration when comparing DFT-calculated $^3J_{CH}$ to those predicted by empirical methods is the time taken to calculate the values. The CPU (central processing unit) time for the NMR calculations varied for the different molecules: strychnine 25.3 hours, camphor 3.0 hours, 2-ethyl-1-indanone (3 conformers) 6.8 hours. The NMR calculations included chemical shifts (as magnetic shielding tensors) for all nuclei and scalar couplings for all pairs of nuclei, not only $^nJ_{CH}$. The geometry optimisation and frequency calculations were not included in these estimates because the molecular structure and conformer populations were also required by the empirical methods. Generating the geometric parameters (such as dihedral angles, bond angles, coupling pathways and substituent patterns) from the Cartesian coordinates of the three molecules (five geometries) was the slowest step, requiring 32.6 seconds of computational time to run two scripts (written in the AWK programming language and MATLAB[168]). The prediction of $^3J_{CH}$ by the fragment-based approach required only 11.2 seconds (205 $^3J_{CH}$) with a MATLAB script, which was not optimised for speed and therefore could be improved further. However, even without optimisation, the total computational time required (43.8 seconds) was only a fraction of the 35.1 hours required for DFT NMR calculations, which is a substantial benefit of empirical approaches to predicting coupling constants.

V.3.3. Limitations and extensions of the fragment-based approach

In comparison to the literature methods, selected for their broad scope, the fragment-based approach showed a more limited applicability to the $^3J_{CH}$ in strychnine, camphor and 2-ethyl-1-indanone. For these compounds the primary weakness of the fragment-based approach is the lack of fragments calculated with nitrogen substituents, the inclusion of a further 13 fragments (HCCC *sp*^{3}-only) with mixed proton, nitrogen and carbon substituents would access an additional 16 $^3J_{CH}$ from strychnine, and therefore improve on the total number of $^3J_{CH}$ calculated by Equation V.2 (Table V.3(b)). A further 33 $^3J_{CH}$ from 27 different fragments would be required to approach the total number of $^3J_{CH}$ that contain 3D structural information, which excludes the 19 $^3J_{CH}$ from aromatic protons. The ease of including new fragments, whether the change is in the substituent pattern or coupling pathway is a benefit of the fragment-based approach.

The fragment-based approach accounts for the effect of two dihedral angles (*Φ* and *Ψ)*, coupling pathway and α, β and γ-substituents, leading to a reduced MAD compared to the literature methods tested, however the SD remained large (>1 Hz), Table V.3. In Figure V.6(d) the (-2.25, -1.75] category showed a higher population (9%) than the corresponding positive (1.75, 2.25] category (6%) and has a significant contribution to the standard deviation. The 9%

population corresponds to five $^3J_{CH}$ where the empirically predicted $^3J_{CH}$ are larger than the experimental values, three $^3J_{CH}$ from H2 to C4, C5 and C7 in camphor (Figure V.1(b)) and two in 2-ethyl-1-indanone, C8-H6b and C10-H6a (Figure V.1(c)). There are several potential causes that might lead to a poor performance of the fragment-based approach for these $^3J_{CH}$, such as the limited performance of methyl groups as proxies for the carbonyls, aromatics and substituted carbons and the presence of multiple coupling pathways (e.g. 4+ bond pathways). The bond angles along the coupling pathway ($\theta_{1\text{-}3}$, Figure V.7) also vary significantly (3-15°) from typical tetrahedral/trigonal planar bond angles due to the nature of the bridged six-membered ring in camphor and the fused aromatic and five-membered ring in 2-ethyl-1-indanone.

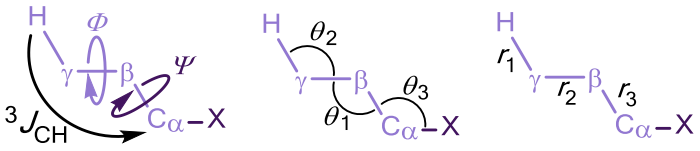


Figure V.7 $^3J_{CH}$ coupling pathway, defining the bond angles ($\theta_{1\text{-}3}$) and bond lengths ($r_{1\text{-}3}$) of interest.

However, other $^3J_{CH}$ empirically predicted by the fragment-based approach show good agreement with the experimentally measured $^3J_{CH}$ while using methyl or OH-groups as proxies for more complex substituents and/or deviations in bond angle. This is highlighted by the empirically predicted $^3J_{CH}$ H1a/b-C3 couplings in camphor, the C3-H1b coupling is 3.3 Hz lower than the experimentally measured value of 5.2 Hz; while for the C3-H1a coupling the difference is smaller (8.8 Hz empirically and 8.6 Hz experimentally) despite the fragment using the same proxies for the substituents. However, as was discussed in detail in section IV.1.1 for carbonyl substituents, the substituent effects depend on the orientation to the coupling pathway and the dihedral angles Φ and Ψ, therefore accounting for the different performance.

V.4. Bond Angle

The effect of bond angle on $^3J_{CH}$ was therefore explored for two of the smallest HCCC fragments, propane (HC001) and butane (HC002), to reduce the computational demands when performing DFT calculations on multiple geometric variations. Geometry optimisation of propane when varying the dihedral angles (Φ and Ψ) in 30° steps led to deviation in bond angles ($\theta_{1\text{-}3}$, Figure V.7) of 2-4° from a tetrahedral value (109.5°) in the geometries calculated: θ_1 = 113.7° ±0.7°, θ_2 = 111.5° ±0.4° and θ_3 = 111.5° ±0.4°. The bond angles in butane with an α-methyl group showed bond angles of θ_1 = 115.1° ±1.5°, θ_2 = 111.6° ±0.5° and θ_3 = 114.5° ±1.5°. Therefore, when considering bond angle as a potential source of error in $^3J_{CH}$ prediction it is important to focus on the deviation in the calculated fragment where geometry optimised bond angles are affected by the presence of substituents. The bond angles were not frozen in the original fragments, for highly substituted fragments freezing the bond angles could

introduce overlap between the substituents and cause the geometry optimisation to fail for certain combinations of Φ and Ψ. Additionally, when applying the fragment-based approach to molecules without strained portions, the geometry optimised bond angles which include substituent effects would be more representative of the typical bond angles in the molecule.

The workflow described in Figure III.5(b) (Chapter III) was used to generate 25,350 geometries and the corresponding $^3J_{CH}$ for propane (HC001) and butane (HC002) by varying Φ and Ψ (0-360° in 30° steps), θ_1 (95-120°), θ_2 and θ_3 (100-120°) in 5° steps. The effect of each of the bond angles is shown in Figure V.8, where the other two bond angles were frozen at 110° for two selected pairs of Φ and Ψ.

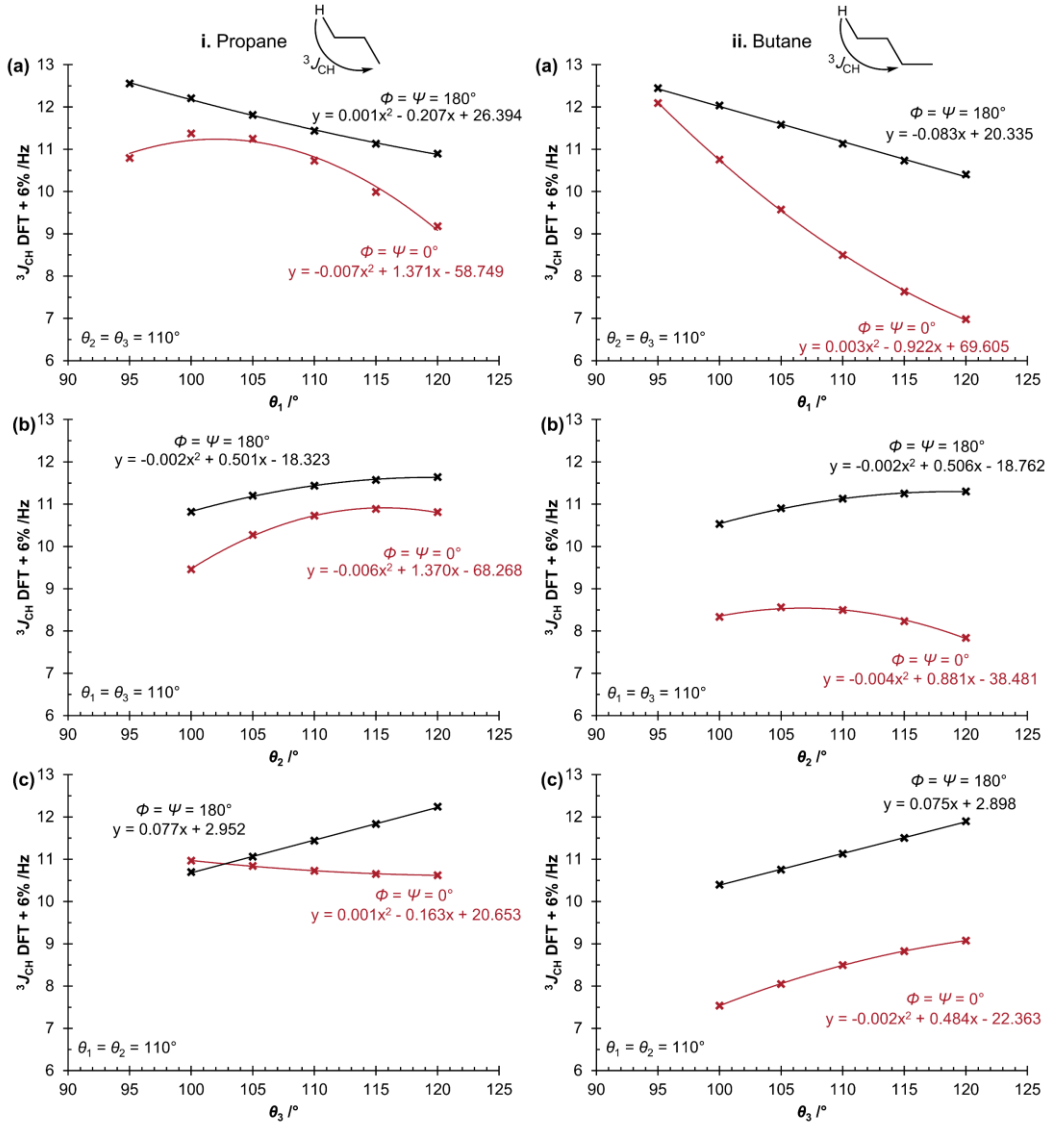


Figure V.8 DFT-calculated $^3J_{CH}$ for i. propane (HC001) and ii. butane (HC002) when varying bond angles (a) θ_1, (b) θ_2 and (c) θ_3.

The relationship between bond angle and $^3J_{CH}$ shows an interaction with the other geometric parameters, for example in Figure V.8i.(c), increasing θ_3 leads to an increase in $^3J_{CH}$ when Φ

= Ψ = 180°, however for Φ = Ψ = 0° a decrease in $^3J_{CH}$ was observed. The effect of each bond angle on $^3J_{CH}$ also differed from propane (HC001, Figure V.8i) on the introduction of an α-methyl group, butane (HC002, Figure V.8ii). Therefore, as was found in the discussion of the carbonyl substituent and the effects of the dihedral angle μ (section IV.1.1), generating a general additive or multiplicative correction factor to improve the accuracy of the fragment-based approach with this data is likely to have limited efficacy due to the complex interactions between the different structural parameters.

An empirical relationship to describe the behaviour of $^3J_{CH}$ when Φ, Ψ and $\theta_{1\text{-}3}$ were varied was investigated for the two fragments. Equation III.1 was limited to the terms necessary to achieve an RMSD ≤0.1 Hz for the fragments without varying bond angle (as described in Table III.2) and then multiplied by $\theta_{1\text{-}3}$ raised to the power of 0-2 as indicated by the mixed linear and quadratic relationships shown in Figure V.8, this resulted in Equation V.4 for propane (HC001) and Equation V.5 for butane (HC002).

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{5}\sum_{j=0}^{3}\sum_{k,l,m=0}^{2} C_{ijklm}\cos(i\Phi)\cos(j\Psi)\,\theta_1^k\theta_2^l\theta_3^m \qquad \text{Equation V.4}$$

Empirical relationship between $^3J_{CH}$, Φ, Ψ and $\theta_{1\text{-}3}$ for propane (HC001).

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{5}\sum_{j=0}^{3}\sum_{k,l,m=0}^{2} C_{ijklm}\cos(i\Phi)\cos(j\Psi)\,\theta_1^k\theta_2^l\theta_3^m + \sum_{i,j=1}^{2}\sum_{k,l,m=0}^{2} S_{ijklm}\sin(i\Phi)\sin(j\Psi)\,\theta_1^k\theta_2^l\theta_3^m \qquad \text{Equation V.5}$$

Empirical relationship between $^3J_{CH}$, Φ, Ψ and $\theta_{1\text{-}3}$ for butane (HC002).

Figure V.9 shows the performance of these fittings, where Equation V.4 for propane (HC001) and Equation V.5 for butane (HC002) resulted in RMSD of 0.2 Hz. Despite the good agreement between the DFT-calculated $^3J_{CH}$ and Equation V.4 and Equation V.5, the performance of more than 50,000 calculations for each fragment to determine the dependence of bond angle requires substantial computational resources (173 days and 328 days of CPU time for propane and butane respectively). When compared to the time taken for molecules similar to strychnine (CPU time 25.3 hours for an NMR calculation), it would only become practical to perform these calculations where a minimum of 164-311 molecular geometries per fragment were under study. Otherwise, given the accuracy of the DFT-calculated $^3J_{CH}$ from the full molecule, it would be more beneficial to not use an empirical approach. Additionally, $^3J_{CH}$ from molecules which show deviation from typical bond angles are likely to be in strained parts of molecules and thus

also show deviation from typical bond lengths. Therefore, accounting for only the bond angle influence with equations like Equation V.4 and Equation V.5 may have a limited improvement in the accuracy of the fragment-based approach.

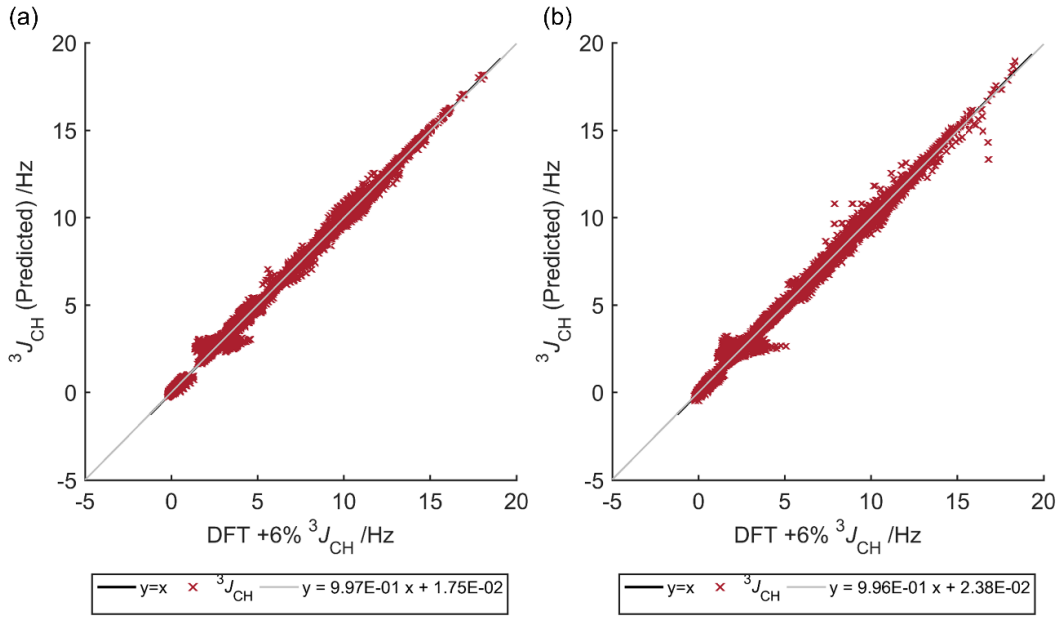


Figure V.9 $^3J_{CH}$ predicted by (a) Equation V.4 for propane (HC001) and (b) Equation V.5 for butane (HC002) compared to DFT-calculated $^3J_{CH}$.

The data calculated for propane and butane with different combinations of Φ, Ψ and $\theta_{1\text{-}3}$ therefore has a limited value in improving the accuracy of the fragment-based approach. However, it does provide estimates of the typical effects of bond angle variation, and therefore an assessment of whether the error in the empirical prediction of a given $^3J_{CH}$ by the fragment-based approach is within what might be expected from this source. Table V.4 shows the maximum effects of θ_1, θ_2 and θ_3 on $^3J_{CH}$ for propane and butane, a variation of 2-5 Hz for a bond angle change of 20-25°. This data suggests a maximum potential change in $^3J_{CH}$ of 0.5 Hz/°, although the average change is only 0.06 Hz/° across the different geometric combinations of Φ and Ψ.

Table V.4 Maximum effect of $\theta_{1\text{-}3}$ on $^3J_{CH}$ for propane (HC001) and butane (HC002).

Maximum effect / Hz	**θ_1 (95-120°)**	**θ_2 (100-120°)**	**θ_3 (100-120°)**
Propane (HC001)	±2.1	±4.9	±3.5
Butane (HC002)	±4.7	±5.2	±5.4

Given that the fragment-based approach gives errors of up to 3 Hz to experimentally measured $^3J_{CH}$ (section V.3.2), the approximate relationship between bond angle and $^3J_{CH}$ from propane/butane suggest a change in the bond angles of 2-17° each. Therefore, the variation

in bond angle of up to 22° observed for camphor it is reasonable to suggest bond angle as a source of error.

For three-bond proton-proton couplings ($^3J_{HH}$) Ōsawa *et al*.[110,111] used an alternative approach to generate bond angle and bond length corrections for $^3J_{HH}$ by using a set of experimentally measured scalar couplings and geometries calculated by molecular mechanics to parameterise a general equation for $^3J_{HH}$ including dihedral angle and substituent electronegativity effects. A similar approach could be applied to $^3J_{CH}$ to generate an additive/multiplicative $\theta_{1\text{-}3}$ correction to $^3J_{CH}$ predicted by the fragment-based approach, however insufficient experimental data has been recorded to both generate and test this method, therefore this approach is returned to in Chapter VI using DFT-calculated $^3J_{CH}$.

V.5. Expanding with literature $^3J_{CH}$

The experimentally measured $^3J_{CH}$ contained only five scalar couplings from the set of 128 fragments containing vinylic carbons in the coupling pathway. Therefore, the dataset was expanded using the experimentally measured $^3J_{CH}$ reported by Parella *et al*.[172] for the HCC=C coupling pathway. It was noted in that work that the typical Karplus-like relationship between $^3J_{CH}$ and Φ was not observed for the HCC=C coupling pathway, in Figure V.10 there is no distinct maximum at Φ = 0/180° or minimum at Φ = 90°. However, the 70 $^3J_{CH}$ reported show 22 different substituent patterns while only four of those substituent patterns show more than four $^3J_{CH}$. Parella *et al*. note that when relating $^3J_{CH}$ from the HCC=C coupling pathway to Φ the study of a closely related structure is required.[172] Therefore, as seen in section IV.2.4, the relationship between $^3J_{CH}$ and Φ may be obscured by substituent effects.

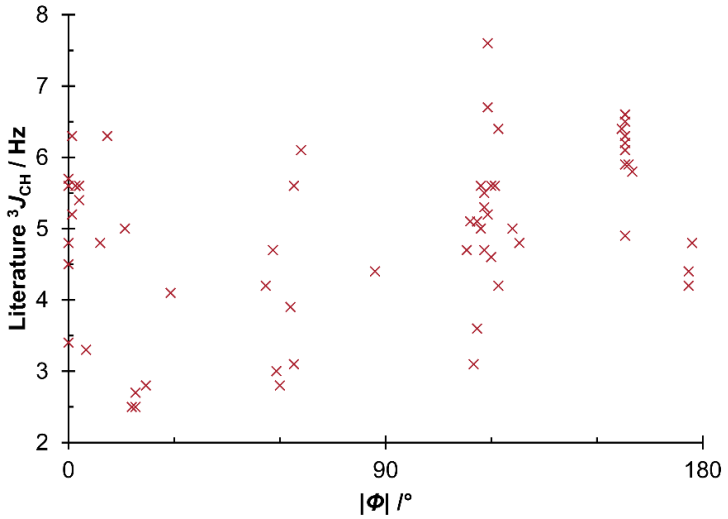


Figure V.10 Experimentally measured $^3J_{CH}$ (±0.3 Hz) correlated with absolute dihedral angle measured from geometries calculated by molecular mechanics, as reported by Parella *et al*. (Table 1).[172]

This set of experimentally measured $^3J_{CH}$ was used to assess the performance of Equation IV.5, which was parameterised using fragments with proton/carbon substituents. The set of 70 experimentally measured $^3J_{CH}$ contained 52 $^3J_{CH}$ with only proton and carbon substituents, however 18 fragments contained oxygen, chlorine and bromine substituents which the fragment-based approach does not include. The dihedral angles were reported as absolute values by Parella *et al*. and therefore the 3D structures were geometry optimised with molecular mechanics (uniform force field, UFF[178]) to estimate the sign of the reported Φ. Figure V.11(a) shows the comparison between the 52 predicted $^3J_{CH}$ and the experimentally measured values. The mean absolute deviation (MAD) of 1.06 Hz was considerably larger than the 0.84 Hz observed for HCCC (sp^3-only) with proton/carbon substituents in section V.3. However, this data includes $^3J_{CH}$ from a number of highly strained molecules including substituted camphenes and norborenes, and the high errors may be attributed to deviation in the bond angles from those used to calculate the fragments. Therefore in Figure V.11(b) the $^3J_{CH}$ were limited to the 18 $^3J_{CH}$ from substituted six-membered rings without bridging groups and showed a significant improvement in both MAD and SD (0.75/0.86 Hz). These substituted six-membered rings have typical bond angles, closer to those for the fragments, suggesting that the improvement in the MAD/SD found is linked to bond angle effects. This further highlights the weakness of empirical methods for predicting $^3J_{CH}$ with complex structural effects (in this case bond angle) beyond the dihedral angle (Φ) between the coupled nuclei.

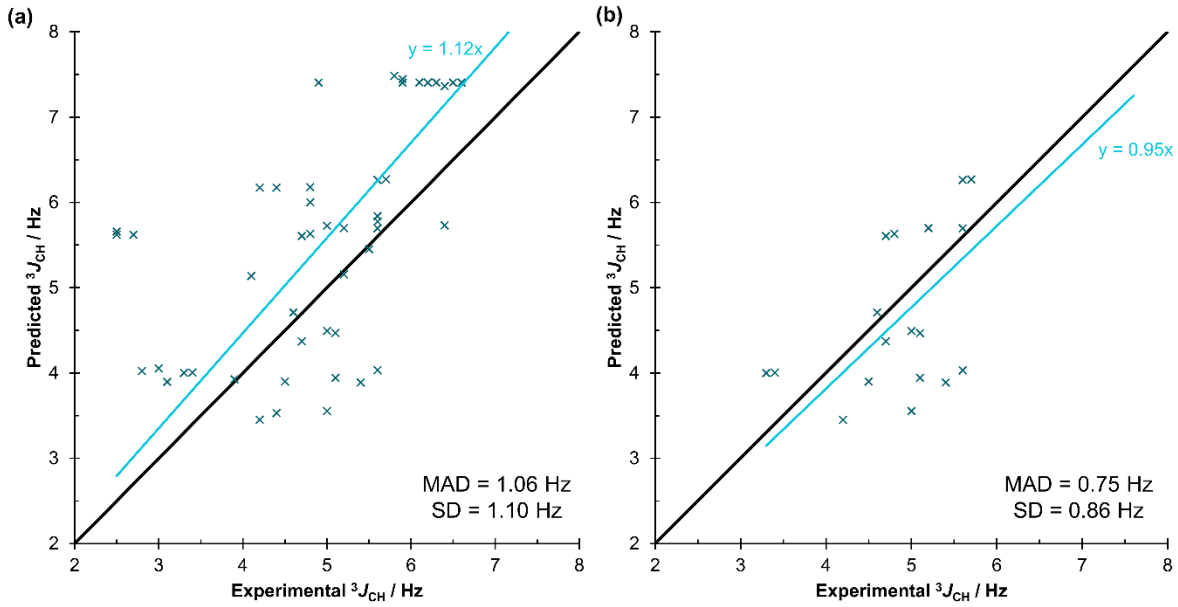


Figure V.11 Comparison of $^3J_{CH}$ predicted by the fragment-based approach to experimentally measured $^3J_{CH}$[172] for (a) the set of 52 $^3J_{CH}$ with proton/carbon substituents (b) the 18 $^3J_{CH}$ from substituted 6-membered rings without bridging groups.

V.6. Summary

This chapter explored the application of the fragment-based approach to empirically predicting $^3J_{CH}$ (as described in Chapter III and Chapter IV) to experimentally measured $^3J_{CH}$ in comparison to the performance of DFT and two empirical methods as described by Aydin *et al*. and Palermo *et al*. The fragment-based approach offered moderate improvements in the prediction of $^3J_{CH}$ for strychnine, camphor and 2-ethyl-1-indanone over the literature equations. However, the application was hindered by the effects of bond angle, multiple coupling pathways and the limited applicability of substituent proxies (such as treating carbonyl substituents as methyl groups). DFT, by far, showed the highest accuracy, however the empirical methods offer a >2800-fold time saving.

The combined effect of bond angle ($\theta_{1\text{-}3}$) and dihedral angle (Φ and Ψ) on $^3J_{CH}$ was explored in detail for propane (HC001) and butane (HC002) with an effect of up to 0.5 Hz/° found. However, to generate similar relationships for >500 fragments was not considered beneficial at this stage; when studying fewer than 164-311 strychnine-like molecules/conformers per fragment the CPU time requirements mean it is preferable to the DFT calculation of the NMR parameters for the full molecular structures.

Even after the inclusion of experimentally measured $^3J_{CH}$ reported by Parella *et al*. only 109 $^3J_{CH}$ were available for comparison to predictions by the fragment-based approach. When separated over the different coupling pathways covered by the fragment-based approach (Table V.5), this means that some substituent variations were not evaluated. This highlights the need for collecting further experimentally measured $^3J_{CH}$ with a known accuracy for rigid, or conformationally simple molecules, a project which is currently underway in collaboration with C4X Discovery as part of an Impact Acceleration Award following this work.

Table V.5 Summary of the number of experimental $^3J_{CH}$ with a comparable value empirically predicted by the fragment-based approach and separated by the subtype of the fragment.

Fragment index	Coupling pathways	Substituents	Number of $^3J_{CH}$
HC001-066 HCO001-192	HCCC (sp^3)	H / C / O	39
HOC001-018 HOCO001-083	HCOC (sp^3)	H / C / O	3
CA001-064	HC(=O)CC HCC(=O)C HCCC(=O)	H / C	10
VIN001-128	HC(=C)CC HCC(=C)C HCCC(=C) HC=CC HCC=C	H / C	57

Chapter VI. Validation of empirical $^3J_{CH}$ prediction against DFT-calculated $^3J_{CH}$

VI.1. Introduction

Chapter V discussed the validation of the fragment-based approach to empirical $^3J_{CH}$ prediction using experimentally measured $^3J_{CH}$ from strychnine, camphor and 2-ethyl-1-indanone. However, these compounds showed limited structural diversity because the experimental dataset was relatively small (102 $^3J_{CH}$) and only explored the accuracy of ~10% of the fragments covered by the fragment-based approach. The DFT-calculated $^3J_{CH}$ from geometry optimised structures were found to be more accurate than the empirical methods in the estimation of experimental $^3J_{CH}$ (Table V.3), therefore empirically predicted $^3J_{CH}$ were compared against a much larger dataset of $^3J_{CH}$ obtained from DFT calculations on molecular structures in the Cambridge Structural Database (CSD)[12]. The CSD contains over 900,000 different small molecule crystal structures, of which 94.6% have 3D coordinates, determined by X-ray or neutron diffraction analysis (for both powders and single crystals)[179]. This database provided a source of 3D molecular structures from which scalar couplings were calculated by DFT to generate a larger set of $^3J_{CH}$ for comparison to empirically predicted $^3J_{CH}$ (Figure VI.1), full details in Chapter VIII.

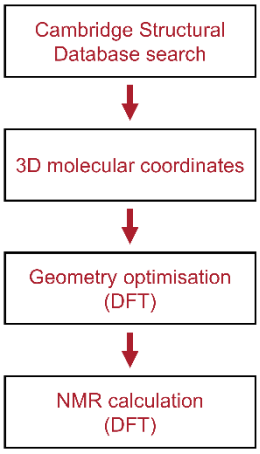
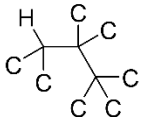


Figure VI.1 Workflow for generating NMR parameters from X-ray crystal structures.

VI.2. Preliminary dataset: Dataset 1



Organics
Not polymeric
No ions
Non-Disordered
No errors
R-Factor ≤0.05
Single crystal structures
Heaviest element: Neon

Figure VI.2 CSD selection criteria for Dataset 1 using ConQuest 1.21[180] and CSD version 5.37[12].

Initially, a small set of 387 molecules was selected from the CSD by searching for organic molecules containing three connected carbons (3° carbon - 4° carbon - 4° carbon) further substituted with carbon (Figure VI.2), which corresponds to fragment index HC064 (Appendix 4). The geometry optimisations and NMR calculations were performed as outlined in Figure VI.1. GaussView5[177] was used to establish the connectivity of the molecules from their Cartesian coordinates and exclude any structures that only contained a solvent molecule or were missing atoms,[181] such as the eight crystal structures which did not explicitly contain proton coordinates (CSD reference codes: BABJUJ, CULDAO, DENPUH, EBIBEW, FAFJUR, KANDUY, LEPYUA, SLSPRM, RULNOB). Any molecules that failed to geometry optimise within 300 hours (real time, full details in Chapter VIII) were also excluded from the dataset. The 387 CSD structures gave a final total of 314 molecules with scalar couplings calculated by DFT, the DFT-calculated $^3J_{CH}$ for these molecules are subsequently referred to as "Dataset 1". The connectivity (including bond orders) for the 314 molecules[181] and the Cartesian coordinates determined from the DFT geometry optimisation were used to predict $^3J_{CH}$ empirically using the workflow described in Chapter V (Figure V.4).

VI.2.1. Empirical prediction of $^3J_{CH}$ for HC064 fragments in Dataset 1

Table VI.1 shows the performance of the empirical prediction of $^3J_{CH}$ compared to the DFT-calculated $^3J_{CH}$ for the 686 scalar couplings from Dataset 1 that match the HC064 fragment, for both the literature equations (Equation V.2, Aydin *et al*.[123] and Equation V.3, Palermo *et al*.[10] and the fragment-based approach described in Chapter III. In comparison to DFT-calculated $^3J_{CH}$, the prediction of $^3J_{CH}$ from only the dihedral angle between the coupled nuclei (Φ) by Equation V.2 (Table VI.1(a)) was the least accurate of the empirical methods tested (MAD/SD 1.34/1.56 Hz). The generalised approach of Palermo *et al*. (Table VI.1(b)) which accounts for the dihedral angle Φ and the effect of β/γ-substituents (electronegativity and orientation) showed the highest accuracy (lowest MAD/SD 0.66/1.07 Hz). The fragment-based approach, which also included the effect of a second dihedral angle (Ψ) and α-substituents, did not further improve the accuracy (MAD/SD 0.73/1.13). The effect of bond angle on $^3J_{CH}$ (as discussed in Chapter V) may contribute to the comparatively poor performance of the fragment-based approach, while the absence of Ψ, α-substituent and bond angle effects in Equation V.3 (Palermo *et al*.) could potentially lead to the cancellation of comparable deviations in $^3J_{CH}$ and a higher apparent accuracy.

Table VI.1 Comparison of empirically predicted to DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for all 686 HC064 fragments in Dataset 1 (a) Aydin *et al*.[123] (b) Palermo *et al*.[10] (c) fragment-based approach.

	(a) Aydin *et al*.	**(b) Palermo *et al*.**	**(c) This work**
Mean absolute deviation (MAD) / Hz	1.34	0.66	0.73
Standard deviation (SD) / Hz	1.56	1.07	1.13

In the DFT-calculated dataset 62% of the bond angles along the $^3J_{CH}$ coupling pathway ($\theta_{1\text{-}3}$) deviate by more than 5° from the average bond angles measured from the isolated HC064 fragment, and 31% deviate by more than 10°. In Chapter V the computational requirements (CPU time) were found to be prohibitive to a fragment-based approach for both dihedral and bond angles effects. However, Dataset 1 provides $^3J_{CH}$ calculated for molecules with a range of dihedral/bond angles, and with bond lengths that are representative of the strained systems that have atypical geometries, unlike the bond lengths in isolated fragments with fixed dihedral/bond angles. Therefore Equation VI.1 was proposed, which introduces a bond angle correction to the $^3J_{CH}$ calculated by the fragment-based approach.

$$^3J_{\mathrm{CH}} = {}^3J_{\mathrm{CH}}(\text{Fragment based approach}) \times (\mathrm{A}_1\Delta\theta_1^2 + \mathrm{A}_2\Delta\theta_2^2 + \mathrm{A}_3\Delta\theta_3^2 + \mathrm{B}_1\Delta\theta_1 + \mathrm{B}_2\Delta\theta_2 + \mathrm{B}_3\Delta\theta_3 + \mathrm{C}) \qquad \text{Equation VI.1}$$

Where $\Delta\theta_n$ is the difference between the mean bond angle ($\bar{\theta}_n$) from the corresponding fragment and the bond angle from the molecule (θ_n): $\Delta\theta_n = \bar{\theta}_n - \theta_n$.

The optimal $A_{1\text{-}3}$, $B_{1\text{-}3}$ and C coefficients were determined by minimising the RMSD between the DFT-calculated $^3J_{CH}$ and the $^3J_{CH}$ predicted using Equation VI.1 (coefficients in Appendix 4). Table VI.2 shows that the introduction of a simple scaling factor (C) had a limited effect on the prediction of $^3J_{CH}$ by Equation VI.1, while the linear bond angle correction ($B_{1\text{-}3}$ and C) resulted in a 0.1 Hz improvement in the MAD/SD (Table VI.2(c)) resulting in a similar accuracy to Palermo *et al*. (Table VI.1(b)). In Chapter V quadratic relationships between bond angle and $^3J_{CH}$ were found, and so a quadratic correction was included ($A_{1\text{-}3}$, $B_{1\text{-}3}$ and C), which gave a substantial improvement in the accuracy of empirically predicted $^3J_{CH}$ in comparison to DFT-calculated $^3J_{CH}$ with MAD <0.5 Hz and SD <1.0 Hz (Table VI.2(d)). However, a larger experimental dataset containing several $^3J_{CH}$ matching the HC064 fragment is required to conclusively test the efficacy of these corrections.

Table VI.2 Comparison of empirically predicted and DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for all HC064 fragments in Dataset 1 (686 $^3J_{CH}$) by the fragment-based approach (a) without or (b-d) with corrections (Equation VI.1) applied.

	No correction	**Bond angle correction: Equation VI.1**		
	(a) Fragment-based approach	(b) Simple scaling factor ($^3J_{CH}$ × C)	(c) Linear ($B_{1\text{-}3}$ & C)	(d) Quadratic ($A_{1\text{-}3}$, $B_{1\text{-}3}$ & C)
Mean absolute deviation (MAD) / Hz	0.73	0.76	0.64	0.46
Standard deviation (SD) / Hz	1.13	1.12	0.98	0.66

VI.2.2. Empirical prediction of $^3J_{CH}$ for all fragments in Dataset 1

Dataset 1 contains a total of 38,625 DFT-calculated $^3J_{CH}$ and therefore provides a large number of $^3J_{CH}$ for exploring the limitations of empirical methods to predict $^3J_{CH}$, including a general bond angle correction. Figure VI.3 shows the performance of empirical $^3J_{CH}$ prediction, where the literature methods were applied using the same limitations used for validation with the experimental dataset (Chapter V). For Equation V.2 (Aydin *et al*.) this meant $^3J_{CH}$ prediction was limited to coupling pathways consisting of three carbons connected by single bonds, while Equation V.3 (Palermo *et al*.) required that the γ and β-carbons were sp^3 hybridised, but the hybridisation-state of the α-carbon was not limited. The fragment-based approach predicted $^3J_{CH}$ only for exact matches with the coupling pathway and substituent pattern (1-bond from the coupling pathway) of the >500 fragments calculated.

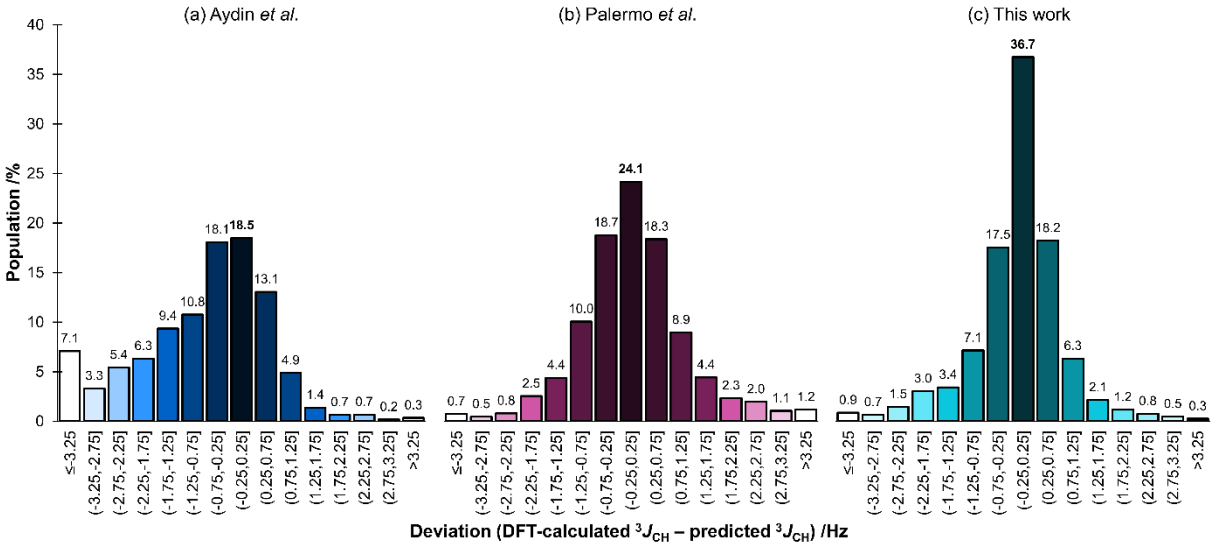


Figure VI.3 Comparison of the deviation between empirically predicted and DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for the full dataset (a) Aydin *et al*.[123] (b) Palermo *et al*.[10] (c) fragment-based approach of this work.

In Table VI.3(a) Aydin *et al*. (Equation V.2), shows lower accuracy than the other two empirical methods and is dominated (Figure VI.3(a)) by a high percentage of outliers with errors of ≤-3.25 Hz. This suggests the equation has a limited use for 3D structural determination despite the slightly broader applicability, predicting 2.9% more $^3J_{CH}$ than the next best method. Further limiting the $^3J_{CH}$ dataset to coupling pathways containing sp^3-only carbons and proton/carbon substituents gave a set of $^3J_{CH}$ that were closer to the experimental dataset used by Aydin *et al*. to determine the coefficients in Equation V.2. The MAD/SD for this reduced set of 25,981 $^3J_{CH}$ were 0.1-0.2 Hz lower than the full dataset, corroborating the findings from comparing empirical prediction of $^3J_{CH}$ to experimentally determined values (Chapter V): that Equation V.2 performs better when the test data closely matches the fitting data. Therefore Dataset 1, which includes a range of compounds with carbon, nitrogen, oxygen and fluorine substituents can be used to fit the coefficients in Equation V.2 and might improve the prediction of $^3J_{CH}$ (sections VI.2.4 and 0).

Equation V.3 (Palermo *et al*.) predicted the 686 DFT-calculated $^3J_{CH}$ from the HC064 fragments in Dataset 1 (Table VI.1) more accurately than the fragment-based approach that included additional structural parameters (Ψ and α-substituents). However, in Table VI.3 the fragment-based approach (c) shows an improvement (0.1-0.2 Hz) compared to that of Palermo *et al*. (b) for all the fragments in Dataset 1. The fragment-based approach also showed a substantial improvement in the accuracy of $^3J_{CH}$ prediction relative to that of Aydin *et al*. (a) with ~0.5 Hz decrease in the SD and approximately half the MAD (0.63 Hz).

Table VI.3 Comparison of empirically predicted and DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for the full dataset (a) Aydin *et al*.[123] (b) Palermo *et al*.[10] (c) fragment-based approach of this work. The numbers in brackets were calculated by limiting the $^3J_{CH}$ to HCCC coupling pathways with sp^3-only carbons and proton/carbon substituents.

	(a) Aydin *et al*.	**(b) Palermo *et al*.**	**(c) This work**
Mean absolute deviation (MAD) / Hz	1.18 (1.07)	0.81 (0.76)	0.63 (0.64)
Standard deviation (SD) / Hz	1.45 (1.29)	1.13 (1.05)	0.95 (0.95)
Total number of predicted $^3J_{CH}$	34,836 (25,981)	33,737 (25,981)	33,128 (25,981)

VI.2.3. Empirical bond angle correction to $^3J_{CH}$ from Dataset 1

The 686 $^3J_{CH}$ that matched the HC064 fragment included geometries with bond angles that deviated from the average values for the HC064 fragment; the prediction of these $^3J_{CH}$ by the fragment-based approach was substantially improved by including a quadratic bond angle correction (Equation VI.1). The 33,128 DFT-calculated $^3J_{CH}$ in Dataset 1 (which can be

predicted by the fragment-based approach) have 66% of the bond angles along the $^3J_{CH}$ coupling pathway deviate by more than 5° from the average bond angles measured from the isolated fragments, and 18% deviate by more than 10°. Therefore, the possibility of using Dataset 1 to generate a general bond angle correction was explored, using the $^3J_{CH}$ calculated for molecules with a range of dihedral/bond angles, and with bond lengths that are representative of those found in the strained systems that lead to atypical geometries. Therefore, the optimal $A_{1\text{-}3}$, $B_{1\text{-}3}$ and C coefficients were determined by minimising the RMSD between the DFT-calculated $^3J_{CH}$ and the $^3J_{CH}$ predicted using Equation VI.1 (Appendix 1).

Table VI.4 shows that the introduction of bond angle corrections led to a moderate improvement in the MAD/SD for the $^3J_{CH}$ calculated by the fragment-based approach. However, unlike the prediction of $^3J_{CH}$ for HC064 fragments there was no significant benefit to the introduction of $\Delta\theta_n^2$ terms. In Chapter V, the full calculations for propane and butane by varying Φ, Ψ and $\theta_{1\text{-}3}$ showed that even a relatively small change (an additional α-methyl group) altered the relationship between $^3J_{CH}$ and the geometric parameters, for example changing from increasing to decreasing $^3J_{CH}$ with increasing θ. Therefore, a specific (rather than general) bond angle correction (such as that in section VI.2.1 for the HC064 fragment) that includes substituent and coupling pathway effects might lead to improved accuracy compared to both DFT-calculated $^3J_{CH}$ and experimental $^3J_{CH}$ for the fragment-based approach. However, the current datasets have two practical limitations meaning this cannot be tested: firstly, Dataset 1 covers only 405 fragments, with ~31% of those fragments with fewer than five DFT-calculated $^3J_{CH}$. Secondly, of the 51 fragments covered by the experimental data there is only an average number of one $^3J_{CH}$ per fragment. Therefore, the current datasets are insufficient both to produce and to conclusively test any benefits of specific bond angle corrections.

Table VI.4 Comparison of 33,128 empirically predicted and DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for all fragments in Dataset 1 (a) no correction, (b) simple scaling factor (Equation VI.1, $A_{1\text{-}3} = B_{1\text{-}3} = 0$), (c) linear bond angle correction (Equation VI.1, $A_{1\text{-}3} = 0$), (d) quadratic bond angle correction (Equation VI.1). The numbers in brackets were calculated by limiting to the 25,981 $^3J_{CH}$ with HCCC coupling pathways with sp^3-only carbons and proton/carbon substituents.

	No correction	**Bond angle correction: Equation VI.1**		
	(a) Fragment-based approach	(b) Simple scaling factor ($^3J_{CH}$ × C)*	(c) Linear ($B_{1\text{-}3}$ & C)	(d) Quadratic ($A_{1\text{-}3}$, $B_{1\text{-}3}$ & C)
Mean absolute deviation (MAD) / Hz	0.63 (0.64)	0.63 (0.64)	0.58 (0.58)	0.59 (0.59)
Standard deviation (SD) / Hz	0.95 (0.95)	0.93 (0.93)	0.87 (0.86)	0.86 (0.86)

***** Full dataset $C = 0.95$, numbers in brackets $C = 0.96$.

In the absence of the data necessary to form a set of specific bond angle corrections, the general correction factor derived from Dataset 1 was tested against the experimental dataset for strychnine, camphor and 2-ethyl-1-indanone (Table VI.5). When considering the full experimental dataset (with no limitations to the fragments tested, Table VI.5(a)), the linear bond angle correction (Table VI.5(b)) showed a small improvement in the MAD and a more significant ~0.1 Hz improvement in the SD.

Therefore, applying the fragment-based approach with the linear bond angle correction (Table VI.5(b), MAD/SD 0.83/1.02 Hz) results in $^3J_{CH}$ prediction with the lowest MAD/SD of the empirical methods tested; Palermo *et al*. previously had the lowest SD against the experimental $^3J_{CH}$ (Table V.3, MAD/SD 0.95/1.06 Hz). The improved performance of the bond angle corrected fragment-based approach therefore justifies the additional requirements in structural interpretation (determining the full substituent pattern, Ψ, $\theta_{1\text{-}3}$).

Table VI.5 Comparison of empirically predicted and experimentally measured $^3J_{CH}$ for strychnine, camphor, and 2-ethyl-1-indanone by (a/c) fragment-based approach and (b/d) linear bond angle correction (Equation VI.1) fitted with matching fragment limitations on $^3J_{CH}$.

	No fragment limitations		**HCCC coupling pathway with *sp*^{3}-only carbons and proton/carbon substituents**	
	(a) Fragment-based approach	(b) + Linear bond angle correction	(c) Fragment-based approach	(d) + Linear bond angle correction
Mean absolute deviation (MAD) / Hz	0.88 (0.93)*	0.83 (0.85)*	0.84	0.80
Standard deviation (SD) / Hz	1.15 (1.23)*	1.02 (1.09)*	1.15	1.03
Total number of predicted and experimental $^3J_{CH}$	53 (39)*	53 (39)*	33	33

* $^3J_{CH}$ where $|\Delta\theta_n| > 5°$ for any n = 1-3 or $\Sigma|\Delta\theta_n| > 10°$.

The molecules within Dataset 1 were chosen by including the HC064 fragment in the CSD selection criteria; this led to a dataset which was biased towards molecules containing a bornane-like substructure (Figure VI.4). Of the initial 387 structures, 125 molecules contain the bornane substructure (two examples are given in Figure VI.4). The experimental data used to test the bond angle correction contains $^3J_{CH}$ from camphor; the close structural relationship between bornane and camphor may bias the bond angle correction factors to perform better for the experimental data than a correction derived from an unbiased dataset (although camphor is not directly included within Dataset 1). Therefore, the accuracy determined for the fragment-based approach with bond angle correction may not be general to a broader range of compounds. In section VI.3 the performance of bond angle correction factors from a dataset which is not biased towards molecules contained in the experimental dataset is investigated (Dataset 2).

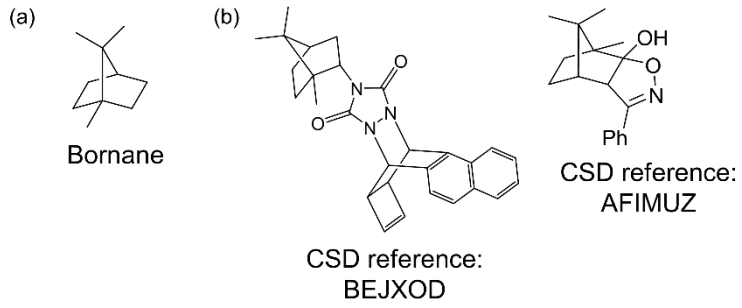


Figure VI.4 (a) Bornane (b) examples of substructure matches from Dataset 1.

VI.2.4. Optimising the performance of Aydin *et al.* (Equation V.2) with Dataset 1

In comparison to the DFT-calculated $^3J_{CH}$ in Table VI.3(a), Aydin *et al*. (Equation V.2) had the broadest applicability of the three empirical approaches, predicting 2.9% more $^3J_{CH}$ than the next best method. However, it showed a limited use for 3D structural determination due to low accuracy, dominated (Figure VI.3(a)) by a high percentage of outliers with errors of ≤-3.25 Hz. It was noted that the accuracy of Equation V.2 improved when the $^3J_{CH}$ were limited to coupling pathways that more closely matched the data used by Aydin *et al*. to determine the coefficients in Equation V.2 (HCCC with *sp*^{3}-only carbons and proton/carbon substituents). Therefore, the equation was fitted using Dataset 1 by minimising the RMSD between the empirically predicted and DFT-calculated $^3J_{CH}$, resulting in Equation VI.2:

$$^3J_{\mathrm{CH}} = 3.70 - 0.24\cos\Phi + 3.48\cos 2\Phi \qquad \text{Equation VI.2}$$

Figure VI.5 shows the improved distribution of errors for Equation VI.2, reducing the systematic overestimation of $^3J_{CH}$ and removing the excess of very negative (≤-3.25 Hz) deviations.

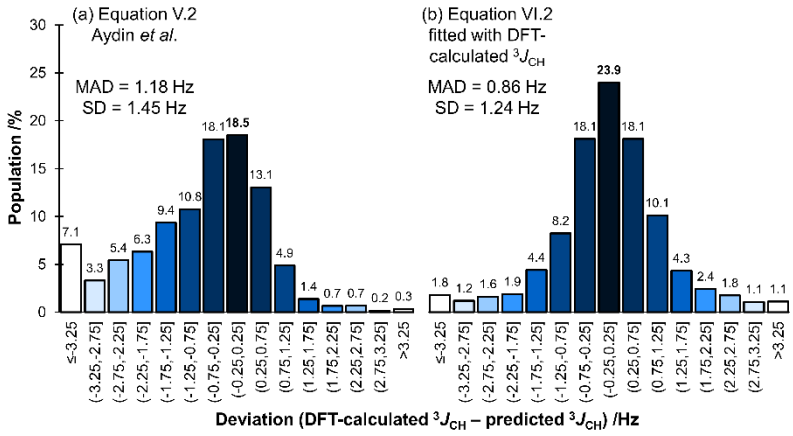


Figure VI.5 Comparison of empirically predicted and DFT-calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p)) for all fragments (a) Equation V.2 Aydin *et al*.[123] (b) Equation VI.2 fitted using the DFT-calculated $^3J_{CH}$.

The performance of the optimised equation is summarised by the improvement in the MAD/SD from 1.18/1.45 Hz(Equation V.2) to 0.86/1.24 Hz(Equation VI.2). This higher accuracy approaches that of Palermo *et al*. MAD/SD 0.81/1.13 Hz (Table VI.3) without the complication

of directly accounting for β/γ-substituent electronegativity. Repeating the optimisation by minimising the RMSD for the 25,981 DFT-calculated $^3J_{CH}$ that form the limited HCCC coupling pathway with *sp*^{3} only carbons and proton/carbon substituents gave Equation VI.3. The MAD/SD decreased from 1.07/1.29 Hz (Equation V.2) to 0.76/1.05 Hz (Equation VI.3), which matches the performance of Palermo *et al*. (Table VI.3), despite not directly accounting for any substituent effects.

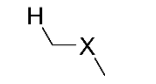
$$^3J_{\mathrm{CH}} = 3.79 - 0.30\cos\Phi + 3.43\cos 2\Phi \qquad \text{Equation VI.3}$$

However, despite the improved performance in the empirical prediction of the DFT-calculated $^3J_{CH}$ used to generate the coefficients for Equation VI.2 and Equation VI.3, the practical use of the equations depends on the performance against experimental data. In Table VI.6 the equations are validated against the experimental data measured for strychnine, camphor and 2-ethyl-1-indanone. A 0.1-0.2 Hz improvement in the MAD and a 0.2-0.4 Hz improvement in the SD was observed for the optimised equations, meaning that the simple *Φ*-only equation achieved the same accuracy as the more complex equation of Palermo *et al*. (0.95/1.06 Hz and 0.99/1.17 Hz, Table V.3) but with the prediction of ~3% more $^3J_{CH}$. The performance of these optimised equations (as posited for the bond angle correction) might, in part, be due to the bias of Dataset 1 towards molecules with a strong structural similarity to part of the experimental dataset. Therefore, the analysis and optimisation are repeated in section VI.3 for the Dataset 2, which shows reduced bias toward the bornane substructure.

Table VI.6 Comparison of empirically predicted and experimentally measured $^3J_{CH}$ for strychnine, camphor and 2-ethyl-1-indanone (a) Aydin *et al*.[123] (b) Equation V.2 fitted with matching limitations on the $^3J_{CH}$ resulting in Equation VI.2 and Equation VI.3.

	H-C-C-C coupling pathway (62 $^3J_{CH}$)		**HCCC coupling pathway with *sp*^{3}-only carbons and proton/carbon substituents (33 $^3J_{CH}$)**	
	(a) Equation V.2 Aydin *et al*.	(b) Equation VI.2	(a) Equation V.2 Aydin *et al*.	(b) Equation VI.3
Mean absolute deviation (MAD) / Hz	1.14	0.90	1.03	0.93
Standard deviation (SD) / Hz	1.38	1.08	1.32	1.11

VI.3. HCXC fragments: Dataset 2



Organics
Not polymeric
No ions
Non-Disordered
No errors
R-Factor ≤0.05
Single crystal structures

Figure VI.6 CSD selection criteria for Dataset 2 using ConQuest 1.21[180] and CSD version 5.38[12].

To establish a set of organic molecules with a reduced bias towards bornane substructures the CSD was searched for molecules containing a proton and carbon connected by three bonds with a carbon, nitrogen or oxygen β-atom (Figure VI.6). This resulted in 109,472 crystal structures, therefore the structures were randomly sampled giving 1,002 structures. The molecules were further divided into three sets of 334 molecules and the geometry optimisations and NMR calculations were performed as outlined in Figure VI.1 for each set by Claire Dickson, William Gerrard[182] and Emily Johnston[183]. The basis set used for the DFT calculations (section VI.3.1) was applicable to elements with atomic numbers lower than krypton, therefore molecules containing heavier elements and any that failed to geometry optimise within 300 hours (real time, full details in Chapter VIII) were excluded from the dataset, giving 860 molecules.

A MATLAB[168] script was developed to determine the connectivity and estimate the bond orders directly from the geometry optimised Cartesian coordinates (full details in Chapter VIII). The script generated the connectivity and estimated the bond orders (primarily based on the interatomic distances) for 462 molecules. 398 molecules failed this process due to a bond order check for each atom in the molecule, which confirms that the total bond order is the same as the expected value for the atom type, for example 1 for proton, 4 for carbon. This eliminated molecules with missing atoms and errors in the connectivity/bond order, however the check also led to a high failure rate for molecules with atoms with multiple potential total bond orders such as sulfur, or delocalisation such as aromatic rings. An additional correction was added to improve the performance for benzene rings, however the diversity of aromatic heterocycles meant that it was challenging to estimate bond order only from connected atoms.

The initial 1,002 CSD structures gave a final set of 462 molecules with scalar couplings calculated by DFT, the DFT-calculated $^{3}J_{CH}$ for these molecules are subsequently referred to as the “Dataset 2”. The connectivity (including bond orders) determined and the Cartesian coordinates determined from the DFT geometry optimisation were used to predict $^{3}J_{CH}$ empirically using the workflow described in Chapter V (Figure V.4).

Dataset 2 covers 428 fragments (more than the 405 fragments in Dataset 1), however the number of fragments with fewer than five DFT-calculated $^{3}J_{CH}$ increased to ~37%. Dataset 1 also contained ~15,000 more $^{3}J_{CH}$ than Dataset 2, despite containing 148 fewer molecules reflecting the bias introduced by limiting the molecules to those containing a connected tertiary and two quaternary carbons, each with carbon substituents.

VI.3.1. Aside: Choosing an optimal DFT method

Dataset 2 was generated for testing empirical approaches to $^{3}J_{CH}$ prediction with reduced structural bias in the molecules towards the experimental test data. However, the dataset also has other applications (Chapter VII) and access to high-performance computing time became available in Bristol[184], so the DFT-calculation method was reinvestigated (Table VI.7) to ensure accurate prediction of long-range scalar couplings ($^{n>1}J_{XY}$). Given the range of molecular size in the CSD, this meant the CPU times for the NMR calculations were also assessed to ensure the method was appropriate for the majority of CSD molecules.

The NMR calculations were performed for strychnine, the largest molecule for which experimental data ($^{n}J_{HH}$ and $^{n}J_{CH}$) were available, including $^{3}J_{CC,}$ $^{1}J_{CC}$[185] and $^{2}J_{CC}$[186]. Three DFT functionals, frequently used for NMR calculations[99,187] were tested, the hybrid functionals B3LYP[80,81] and mPW1PW91[82] and the long-range corrected ωB97X-D[83,84]. These functionals were combined with twelve different basis sets, five Pople basis sets (3-21G,[61] 6-31G,[62-68] 6-31G(d), 6-311G(d,p)[69,70] and 6-311++G(d,p)); one Pople basis set optimised for scalar coupling calculation (6-311++G(d,p)-*J*)[75]; four correlation consistent basis sets (cc-pVDZ, cc-pVTZ, aug-cc-pVDZ and aug-cc-pVTZ[188-190]); and the universal Gaussian basis set (UGBS [70,191]). The NMR calculations for each combination of functional and basis set were performed for the same geometry of strychnine. The geometry optimisation was performed with mPW1PW91/6-311G(d,p) and the IEFPCM solvent model for chloroform, (full details in Chapter VIII, Experimental).

Table VI.7 Comparison of the RMSD between experimental (strychnine) and DFT-calculated $^{n}J_{XY}$ for different functional, basis set and mixed combinations. The darker the red shading the higher the RMSD, the lowest RMSD are indicted by a **bold** font.

DFT method			CPU time /hours	RMSD / Hz						
Functional	Basis Set	Mixed		$^{n>1}J_{XY}$	$^{n}J_{HH}$	$^{n>1}J_{CH}$	$^{n>1}J_{CC}$	$^{1}J_{XY}$	$^{1}J_{CH}$	$^{1}J_{CC}$
ωB97X-D	6-311++G **-J	No	203	**0.57**	0.73	0.49	0.57	4.4	5.9	1.7
ωB97X-D	UGBS	No	2646	0.58	0.77	0.49	0.51	5.0	7.0	1.5
ωB97X-D	6-311++G (d,p)	Yes	180	0.59	0.72	0.54	0.46	7.5	10.3	2.3
ωB97X-D	6-311G (d,p)	Yes	52	0.59	0.72	0.56	0.46	7.7	10.6	2.2
ωB97X-D	cc-pVTZ	Yes	349	0.59	0.74	0.54	0.51	7.9	10.9	2.3
ωB97X-D	6-31G (d)	Yes	30	0.64	**0.71**	0.63	0.55	9.3	12.8	3.1
ωB97X-D	6-31G	Yes	16	0.65	0.87	0.56	0.46	7.0	9.6	2.2
mPW1PW91	6-311++G (d,p)	No	87	0.67	0.81	0.63	0.47	9.6	13.5	1.9
ωB97X-D	cc-pVDZ	Yes	60	0.67	0.80	0.65	0.47	9.3	13.0	1.7
mPW1PW91	cc-pVDZ	Yes	51	0.68	1.05	**0.48**	0.49	3.1	4.2	1.4
mPW1PW91	6-311G (d,p)	No	32	0.68	0.82	0.64	0.51	8.6	11.9	2.3
ωB97X-D	3-21G	Yes	11	0.75	0.83	0.75	0.46	12.4	16.5	6.0
mPW1PW91	cc-pVTZ	Yes	344	0.75	1.21	0.50	0.46	2.5	3.1	1.6
mPW1PW91	6-311G (d,p)	Yes	47	0.76	1.24	0.49	0.47	2.7	3.3	1.7
B3LYP	cc-pVDZ	Yes	51	0.76	0.98	0.70	0.45	3.1	3.6	2.5
B3LYP	aug-cc-pVDZ	Yes	456	0.76	1.02	0.67	0.43	2.9	3.6	1.8
mPW1PW91	6-311++G (d,p)	Yes	159	0.77	1.28	0.49	0.45	2.6	3.1	1.8
B3LYP	6-311G (d,p)	Yes	46	0.79	1.10	0.67	0.46	3.4	4.6	1.4
B3LYP	6-311G (d,p)	No	31	0.81	0.98	0.78	0.41	5.1	6.3	3.6
B3LYP	cc-pVTZ	Yes	323	0.81	1.09	0.71	0.52	3.5	4.7	1.5
mPW1PW91	6-31G (d)	Yes	28	0.82	1.39	0.49	0.41	3.8	4.8	2.2
B3LYP	6-311++G (d,p)	Yes	155	0.83	1.14	0.70	0.52	2.4	3.1	1.3
B3LYP	6-31G (d)	Yes	28	0.83	1.14	0.70	0.52	2.4	3.1	1.3
B3LYP	6-311++G (d,p)	No	101	0.83	1.03	0.78	**0.40**	5.6	7.6	2.2
B3LYP	3-21G	Yes	10	0.87	1.35	0.64	0.47	5.3	5.2	5.4
mPW1PW91	aug-cc-pVTZ	No	2760	0.87	1.10	0.81	0.57	16.8	21.4	10.5
B3LYP	6-311++G **-J	No	234	0.87	1.26	0.71	0.54	6.9	9.4	2.9
mPW1PW91	6-311++G **-J	No	173	0.88	1.40	0.60	0.59	2.2	2.6	1.6
mPW1PW91	cc-pVTZ	No	264	0.89	0.88	0.90	0.88	13.8	18.1	7.3
mPW1PW91	3-21G	Yes	8	0.94	1.54	0.60	0.58	9.5	11.6	6.8
B3LYP	6-31G	Yes	13	0.94	1.55	0.61	0.47	4.3	5.9	1.4
ωB97X-D	6-311G (d,p)	No	33	0.97	1.29	0.87	0.44	12.6	17.7	2.1
ωB97X-D	6-311++G (d,p)	No	137	0.97	1.30	0.88	0.42	13.8	19.4	2.1
mPW1PW91	6-31G	Yes	14	1.04	1.71	0.67	0.68	2.3	2.7	1.8
B3LYP	cc-pVTZ	No	281	1.07	1.17	1.05	0.89	9.7	11.9	6.7
mPW1PW91	UGBS	No	2492	1.10	1.75	0.74	0.77	**1.8**	**2.2**	**1.2**
ωB97X-D	cc-pVTZ	No	269	1.22	1.39	1.19	0.96	18.9	25.7	7.4
ωB97X-D	6-31G	No	6	1.35	1.66	1.12	1.78	16.0	14.5	17.3
B3LYP	6-31G	No	5	1.53	2.20	1.11	1.73	21.3	23.5	19.0
B3LYP	6-31G (d)	No	12	1.60	0.93	1.57	2.74	14.0	12.2	15.6
mPW1PW91	6-31G (d)	No	12	1.62	1.69	1.33	2.71	13.3	12.1	14.3
ωB97X-D	6-31G (d)	No	15	1.66	0.92	1.68	2.72	14.4	17.1	11.0
mPW1PW91	3-21G	No	4	1.67	1.38	1.69	2.13	21.7	30.6	2.8
B3LYP	3-21G	No	4	1.71	1.59	1.74	1.76	21.5	30.3	3.4
mPW1PW91	cc-pVDZ	No	22	1.75	1.77	1.67	2.14	22.4	13.2	28.8
ωB97X-D	aug-cc-pVDZ	No	422*	1.76	1.94	1.56	2.39	42.0	7.7	58.9
mPW1PW91	aug-cc-pVDZ	No	201	1.78	1.36	1.78	2.55	43.8	9.0	61.2
ωB97X-D	3-21G	No	5	1.94	2.08	1.89	1.81	26.3	37.0	4.3
B3LYP	cc-pVDZ	No	21	2.06	2.17	1.97	2.34	22.3	13.6	28.5
ωB97X-D	cc-pVDZ	No	23	2.16	2.42	2.02	2.28	19.9	9.9	26.3
mPW1PW91	6-31G	No	5	2.19	3.12	1.61	2.45	21.4	22.7	20.0

*Additional convergence criteria: integral = ultrafine and convergence = 10.

The “mixed” method in Gaussian09[78] was introduced because it improves the basis set description of electron density at the nucleus for valence-oriented basis sets and therefore improves the accuracy of calculated scalar couplings[79]. Firstly, the Fermi contact contribution to the coupling is calculated with the basis set uncontracted and additional tight *s* functions, then the contracted (unmodified) basis set is used to calculate the spin dipolar, paramagnetic spin-orbit and diamagnetic spin-orbit contributions to the coupling. The mixed method was originally tested against the uncontracted UGBS2P basis set with B3LYP for one and two-bond scalar couplings in very small molecules (<7 atoms)[79], therefore in Table VI.7 the method was investigated for longer range couplings and with multiple functionals against experimental data.

The data for strychnine in Table VI.7 includes the RMSD for long range ($^{n>1}J$) and one-bond (^{1}J) scalar couplings with the CPU time (full details in Chapter VIII) for each combination of functional, basis set and with/without mixed. Some combinations of functional, basis set and mixed are absent from Table VI.7 either due to a failure to converge or the calculation requiring more than 300 hours (real time). The combination of ωB97X-D with 6-311g (d,p) and the mixed basis set (underlined in Table VI.7) offered the highest accuracy with <100 hours of CPU time for strychnine. Using the number of carbons as a marker of the molecules size (strychnine 21 carbons) the 109,472 crystal structures in the initial CSD search show an average of 19 carbons per structure (up to 156 carbons). Therefore, methods that perform the NMR calculation with <100 hours for strychnine should cover a reasonable fraction of Dataset 2 without being limited by computational resources. The “mixed” method with the ωB97X-D functional and 6-311G (d,p) basis set offered a reasonable balance of $^{n>1}J_{XY}$ accuracy against the computational resources required (underlined in Table VI.7), therefore this was used for the NMR calculations for the molecules in Dataset 2.

VI.3.2. Empirical prediction of $^{3}J_{CH}$ for Dataset 2

Dataset 2 contains 23,398 DFT-calculated $^{3}J_{CH}$ from 462 molecules with reduced bias towards the experimental dataset (than Dataset 1) for exploring the limitations of empirical methods to predict $^{3}J_{CH}$ and the production of a general bond angle correction. Figure VI.7 shows the performance of empirical $^{3}J_{CH}$ prediction, where the literature methods were applied using the same limitations used for validation with the experimental and DFT datasets.

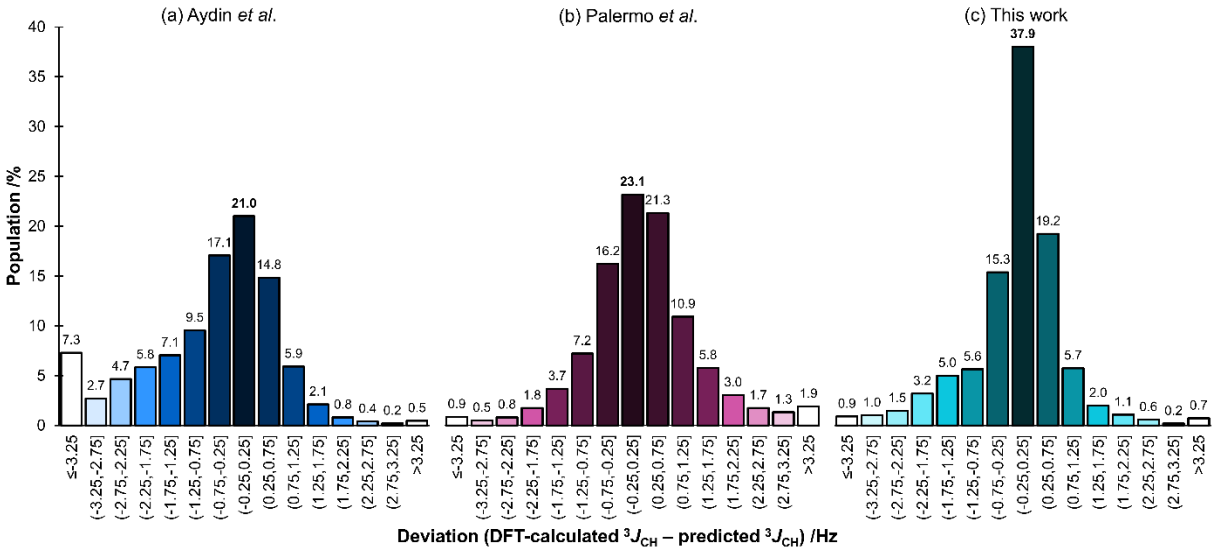


Figure VI.7 Comparison of the deviation between empirically predicted and DFT-calculated $^3J_{CH}$ (ωB97X-D/6-311G(d,p)) for the full dataset (a) Aydin *et al*.[123] (b) Palermo *et al*.[10] (c) fragment-based approach of this work.

In Table VI.8(a), Aydin *et al*. (Equation V.2) showed the lowest accuracy of the empirical methods and is dominated (Figure VI.7(a)) by a high percentage of outliers with errors of ≤-3.25 Hz, similar to Dataset 1 (Figure VI.3(a)). Despite the broad applicability (~4% more $^3J_{CH}$ predicted than the next best method), the low accuracy results in a limited use for 3D structural determination. Further limiting the $^3J_{CH}$ dataset to coupling pathways containing sp^3-only carbons and proton/carbon substituents gave a set of $^3J_{CH}$ that were closer to the experimental dataset used by Aydin *et al*. to determine the coefficients in Equation V.2. As observed for the experimental $^3J_{CH}$ and Dataset 1 the MAD/SD for this reduced set of 5,211 $^3J_{CH}$ were 0.1-0.3 Hz lower. Optimising the coefficients using the DFT-calculated $^3J_{CH}$ from Dataset 1 (section VI.2.4) led to further improvements for the prediction of experimental data. Therefore, this is repeated (section VI.3.4) for the Dataset 2, to separate the improvement from the bias in the dataset.

The fragment-based approach (Table VI.8(c)) showed higher accuracy (MAD/SD 0.66/1.04 Hz) in the prediction of DFT-calculated $^3J_{CH}$ than Palermo *et al*. (MAD/SD 0.85/1.22 Hz). The fragment-based approach also showed a substantial improvement in the accuracy of $^3J_{CH}$ prediction relative to that of Aydin *et al*. (a) with ~0.5 Hz decrease in the SD and approximately half the MAD (0.66 Hz). However, despite the higher accuracy of the fragment-based approach, Palermo *et al*. calculated ~10% more and Aydin *et al*. ~14% more $^3J_{CH}$. This reinforces the necessity of calculating further fragments for the fragment-based approach, as the lower applicability is the primary drawback of the method.

Table VI.8 Comparison of empirically predicted and DFT-calculated $^3J_{CH}$ (ωB97X-D/6-311G(d,p)) for the all fragments in Dataset 2 (a) Aydin *et al*.[123] (b) Palermo *et al*.[10] (c) fragment-based approach of this work. The numbers in brackets were calculated by limiting the $^3J_{CH}$ to HCCC coupling pathways with sp^3-only carbons and proton/carbon substituents.

	(a) Aydin *et al*.	**(b) Palermo *et al*.**	**(c) This work**
Mean absolute deviation (MAD) / Hz	1.15 (0.98)	0.85 (0.83)	0.66 (0.68)
Standard deviation (SD) / Hz	1.53 (1.36)	1.22 (1.17)	1.04 (1.08)
Total number of predicted $^3J_{CH}$	12,277 (5,211)	11,301 (5,211)	8,924 (5,211)

VI.3.3. Empirical bond angle correction to $^3J_{CH}$ Dataset 2

The $^3J_{CH}$ predicted by the fragment-based approach for the Dataset 1 showed improved agreement with the DFT-calculated values on the inclusion of a bond angle correction (Equation VI.1). 66% of the bond angles along the $^3J_{CH}$ coupling pathway for Dataset 1 deviated by more than 5° (from the average for the isolated fragments), and 18% deviated by more than 10°. However, for the 8,924 $^3J_{CH}$ in Dataset 2 only 19% deviated by more than 5° and 7% by more than 10°, suggesting that the molecules contain fewer strained portions and more typical bond angles. Dataset 2 was used to generate a general bond angle correction, using the $^3J_{CH}$ calculated for molecules with a range of dihedral/bond angles, and with bond lengths that are representative of those found in the strained systems that lead to atypical geometries but without a strong bias toward the bornane substructure. Therefore, the optimal $A_{1\text{-}3}$, $B_{1\text{-}3}$ and C coefficients were determined by minimising the RMSD between the DFT-calculated $^3J_{CH}$ and the $^3J_{CH}$ predicted using Equation VI.1 (Appendix 1).

Table VI.9 shows that the introduction of quadratic bond angle corrections led to a moderate improvement in the MAD/SD for the $^3J_{CH}$ calculated by the fragment-based approach. However, unlike the prediction of $^3J_{CH}$ for Dataset 2 the linear bond angle correction had a slightly reduced benefit. Dataset 2 (like Dataset 1) is still insufficient to produce specific bond angle corrections (such as that in section VI.2.1 for the HC064 fragment) where the included substituent and coupling pathway effects might lead to improved accuracy compared to both DFT-calculated $^3J_{CH}$ and experimental $^3J_{CH}$ for the fragment-based approach.

Table VI.9 Comparison of 8,924 empirically predicted and DFT-calculated $^3J_{CH}$ (ωB97X-D/6-311G(d,p)) for all fragments. (a) No correction, (b) simple scaling factor (Equation VI.1, $A_{1\text{-}3} = B_{1\text{-}3} = 0$), (c) linear bond angle correction (Equation VI.1, $A_{1\text{-}3} = 0$), (d) quadratic bond angle correction (Equation VI.1). The numbers in brackets were calculated by limiting to 5,211 $^3J_{CH}$ with HCCC coupling pathways with sp^3-only carbons and proton/carbon substituents.

	No correction	**Bond angle correction: Equation VI.1**		
	(a) Fragment-based approach	(b) Simple scaling factor ($^3J_{CH}$ × C)*	(c) Linear ($B_{1\text{-}3}$ & C)	(d) Quadratic ($A_{1\text{-}3}$, $B_{1\text{-}3}$ & C)
Mean absolute deviation (MAD) / Hz	0.66 (0.68)	0.66 (0.68)	0.63 (0.68)	0.61 (0.63)
Standard deviation (SD) / Hz	1.04 (1.08)	1.00 (1.06)	0.97 (1.02)	0.91 (0.92)

***** Full dataset $C = 0.95$, numbers in brackets $C = 0.96$.

The general correction factors derived from Dataset 2 were tested against the experimental dataset for strychnine, camphor and 2-ethyl-1-indanone (Table VI.10). Despite the quadratic bond angle correction to the fragment-based approach performing well for DFT-calculated $^3J_{CH}$ (Table VI.9(d)), it showed no significant benefit in comparison to the experimental data (Table VI.10(c/f)). Instead, when considering the full experimental dataset (with no limitations to the fragments tested, Table VI.10(a-c)), the linear bond angle correction (Table VI.10(b)) showed the largest improvement in the SD (1.15 compared to 1.04 Hz). This reduction in the SD results in an empirical approach to $^3J_{CH}$ prediction with a lower MAD/SD (0.86/1.04 Hz) than Palermo *et al*. (Table V.3, MAD/SD 0.95/1.06 Hz) which previously had the lowest SD against the experimental $^3J_{CH}$.

Table VI.10 Comparison of empirically predicted and experimentally measured $^3J_{CH}$ for strychnine, camphor, and 2-ethyl-1-indanone by (a/d) fragment-based approach and (b/e) linear bond angle correction (c/f) quadratic bond angle correction (Equation VI.1) fitted with matching fragment limitations.

	No limitations (53 $^3J_{CH}$)			**HCCC coupling pathway with sp^3-only carbons and proton/carbon substituents (33 $^3J_{CH}$)**		
	(a) Fragment-based approach	(b) Linear ($B_{1\text{-}3}$ & C)	(c) Quadratic ($A_{1\text{-}3}$, $B_{1\text{-}3}$ & C)	(d) Fragment-based approach	(e) Linear ($B_{1\text{-}3}$ & C)	(f) Quadratic ($A_{1\text{-}3}$, $B_{1\text{-}3}$ & C)
Mean absolute deviation (MAD) / Hz	0.88	0.86	0.88	0.84	0.88	0.93
Standard deviation (SD) / Hz	1.15	1.04	1.07	1.15	1.07	1.16

VI.3.4. Optimising the performance of Aydin *et al*. (Equation V.2) with Dataset 2

In comparison to the DFT-calculated $^3J_{CH}$ in Table VI.8(a), Aydin *et al*. (Equation V.2) had the broadest applicability of the three empirical approaches, predicting 4% more $^3J_{CH}$ than the Palermo *et al*., and 14% more than the fragment-based approach. However, it showed a limited use for 3D structural determination due to low accuracy (in comparison to both experimentally measured and DFT-calculated $^3J_{CH}$), dominated (Figure VI.7(a)) by a high percentage (7%) of outliers with errors of ≤-3.25 Hz. Fitting the equation with Dataset 1 by minimising the RMSD between the empirically predicted and DFT-calculated $^3J_{CH}$, resulting reduced MAD/SD for both the DFT-calculated and experimentally measured $^3J_{CH}$. However, it was suggested that this improvement might be attributed to the bias of the Dataset 1 toward structures related to those in the experimental dataset. Therefore, the equation was fitted using Dataset 2 by minimising the RMSD between the empirically predicted and DFT-calculated $^3J_{CH}$, resulting in Equation VI.4:

$$^3J_{\mathrm{CH}} = 3.80 - 0.72\cos\Phi + 3.07\cos 2\Phi \qquad \text{Equation VI.4}$$

Figure VI.8 shows the improved distribution of errors for the optimised Equation VI.4, reducing the systematic overestimation of $^3J_{CH}$ and removing the excess of very negative (≤-3.25 Hz) deviations.

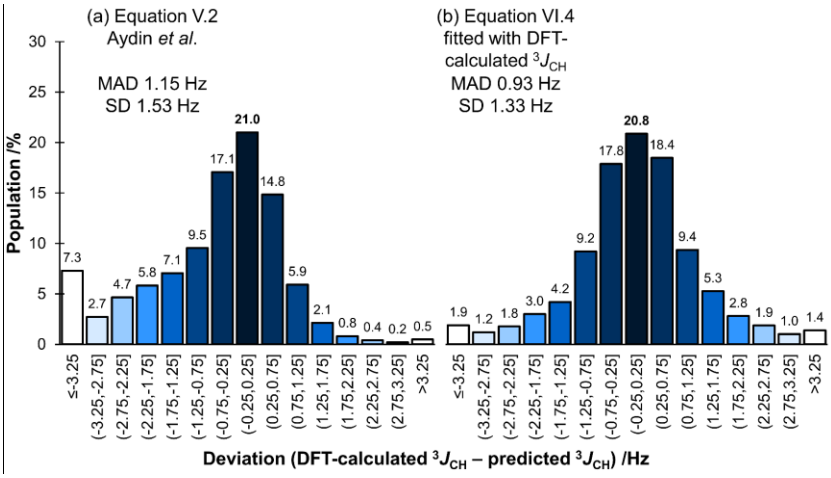


Figure VI.8 Comparison of empirically predicted and DFT-calculated $^3J_{CH}$ (ωB97X-D/6-311G(d,p)) for all fragments (a) Equation V.2 Aydin *et al*.[123] (b) Equation VI.4 fitted using the DFT-calculated $^3J_{CH}$.

The performance of the optimised equation is summarised by the improvement in the MAD/SD from 1.15/1.53 Hz (Equation V.2) to 0.93/1.33 Hz (Equation VI.4). This higher accuracy approaches that of Palermo *et al*. MAD/SD 0.85/1.22 Hz (Table VI.8) without the complication of directly accounting for β/γ-substituent electronegativity. Repeating the optimisation by minimising the RMSD for the 5,211 DFT-calculated $^3J_{CH}$ that form the limited HCCC coupling pathway with sp^3 only carbons and proton/carbon substituents gave Equation VI.5. The

MAD/SD decreased from 0.98/1.36 Hz (Equation V.2) to 0.83/1.18 Hz (Equation VI.5) which also matches the performance of Palermo *et al*. (Table VI.8), despite not directly accounting for any substituent effects.

$$^{3}J_{\mathrm{CH}} = 4.08 - 0.54\cos\Phi + 3.36\cos 2\Phi \qquad \text{Equation VI.5}$$

However, despite the improved performance in the empirical prediction of $^{3}J_{CH}$ against the DFT-calculated $^{3}J_{CH}$ used to generate the coefficients for Equation VI.4 and Equation VI.5, the practical use of the equations depends on the performance against experimental data. For Equation VI.2 and Equation VI.3 (fitted using Dataset 1) it was suggested that the structural similarities between the DFT-calculated and experimental structures might have contributed to the improved prediction of $^{3}J_{CH}$. In Table VI.11 the same validation has been performed for Equation VI.4 and Equation VI.5 against the experimental data measured for strychnine, camphor and 2-ethyl-1-indanone. A 0.1-0.3 Hz improvement in the MAD and a 0.2-0.3 Hz improvement in the SD was observed for the optimised equations, meaning that the simple Φ-only equation achieved the same accuracy as the more complex equation of Palermo *et al*. (0.95/1.06 Hz and 0.99/1.17 Hz, Table V.3) but with the prediction of ~3-4% more $^{3}J_{CH}$. The slightly improved performance of Equation VI.4 (optimised with Dataset 2) suggests that the bias of Dataset 1 was instead limiting the improvement in $^{3}J_{CH}$ prediction rather than contributing to it.

Table VI.11 Comparison of empirically predicted and experimentally measured $^{3}J_{CH}$ for strychnine, camphor and 2-ethyl-1-indanone (a) Aydin *et al*.[123] (b) Equation V.2 fitted with matching limitations on the $^{3}J_{CH}$ resulting in Equation VI.4 and Equation VI.5.

	H-C-C-C coupling pathway (62 $^{3}J_{CH}$)		**HCCC coupling pathway with sp^3-only carbons and proton/carbon substituents (33 $^{3}J_{CH}$)**	
	(a) Equation V.2 Aydin *et al*.	(b) Equation VI.4	(a) Equation V.2 Aydin *et al*.	(b) Equation VI.5
Mean absolute deviation (MAD) / Hz	1.14	0.82	1.03	0.87
Standard deviation (SD) / Hz	1.38	1.08	1.32	1.13

VI.4. Summary

This chapter established two sets of DFT-calculated $^{3}J_{CH}$, one dataset consisting of 38,625 values (Dataset 1) and one 23,398 values (Dataset 2), which covered ~75% of fragments described by the fragment-based approach. The prediction of $^{3}J_{CH}$ by the fragment-based approach was the most accurate (MAD 0.6-0.7 Hz, SD ~1.0 Hz) in comparison to the DFT

datasets. However, it also predicted the fewest number of $^3J_{CH}$ values (4-14% fewer $^3J_{CH}$) of the empirical methods tested.

Datasets 1 and 2 were also used to establish a general bond angle correction to the fragment-based approach (Equation IV.1), reducing the MAD/SD by ~0.1 Hz in comparison to the DFT-calculated $^3J_{CH}$ (Table VI.4 and Table VI.9). The coefficients in Equation V.2 were also optimised (Equation VI.4), significantly improving the accuracy of the simplest equation (only Φ dependent) to predict $^3J_{CH}$ MAD/SD (1.15/1.53 Hz to 0.86/1.24 Hz).

The dataset of experimentally measured $^3J_{CH}$ described in Chapter V was used to test the empirical bond angle corrections and the optimised coefficients for Aydin *et al*. Table VI.12 summarises the performance of the most accurate version of each empirical method tested and the DFT-calculated $^3J_{CH}$ for comparison. This shows that DFT offers the highest accuracy and number of $^3J_{CH}$, however empirical methods offer ~2800-fold time saving in the calculation of $^3J_{CH}$.

Table VI.12 Summary of the most accurate version of the methods explored for the estimation of $^3J_{CH}$, validation against experimentally measured $^3J_{CH}$ from strychnine, camphor and 2-ethyl-1-indanone.

	DFT (mPW1PW91/6-311G(d,p))	**Optimised Aydin *et al*. (Dataset 2)**	**Palermo *et al*.**	**Fragment-based approach with bond angle correction***
Equation		Equation VI.4	Equation V.3	Equation VI.1
Structural information required	Full Cartesian coordinates	Φ	Φ β/γ-substituent orientation and electronegativity	Φ, Ψ, $\theta_{1\text{-}3}$ Coupling pathway α/β/γ-substituent pattern
MAD / Hz	0.37	0.82	0.95	0.86
SD / Hz	0.44	1.08	1.06	1.04
Number of $^3J_{CH}$	102	62	60	53

* Linear correction from Dataset 2

Equation VI.4, which was optimised using Dataset 2, shows the highest number of $^3J_{CH}$ predicted (62) compared to the experimental dataset. However, the inclusion of amine (NH_2) substituents for the saturated HCCC coupling pathway (192 fragments) and a β-nitrogen in the coupling pathway (HCNC, 68 fragments) to the fragment-based approach would give a total of 64 $^3J_{CH}$ for comparison to the experimental data (Appendix 5), surpassing the other empirical methods. Further additions such as an α-aromatic carbon (HCCC), vinylic coupling pathways with oxygen or nitrogen and saturated HCCC coupling pathways with multiple NH_2/OH substituents approach the limit of $^3J_{CH}$ fragments that contain 3D structural information. Strychnine and 2-ethyl-1-indanone contain 16 $^3J_{CH}$ where the α, β and γ carbons are aromatic, for example. This highlights the need for calculating $^3J_{CH}$ for additional fragments, a project which is currently underway in collaboration with C4X Discovery as part of a follow-on Impact Acceleration Award.

Chapter VII. Conclusions and Future Work

This thesis aimed to improve methods used to study the three-dimensional (3D) structures of small molecules using solution-state NMR spectroscopy. It focussed on long-range proton-carbon scalar coupling ($^{n>1}J_{CH}$) and the generation/validation of empirical equations to predict $^{3}J_{CH}$ from molecular structures.

VII.1. Accurate experimental measurement of $^{n>1}J_{CH}$

The measurement of 55 $^{n>1}J_{CH}$ couplings from coupled ^{13}C spectra in Chapter II allowed the assessment of the accuracy of selected 2-dimensional methods for measuring $^{n>1}J_{CH}$ in model compounds (strychnine and camphor).

The IPAP-based analysis of HSQMBC and HMBC data provided largest number of the expected $^{n>1}J_{CH}$ values in a relatively short amount of experimental and analysis time with good accuracy (mean deviations of <0.5 Hz from coupled ^{13}C data). The ^{1}H-selective homonuclear decoupled *J*-scaled F1 evolution of $^{n>1}J_{CH}$ in experiments such as EXSIDE provided the simplest spectra for measurement of $^{n>1}J_{CH}$ but used a substantial amount of experiment time when targeting all possible ^{1}H in strychnine and camphor. The accuracy of raw EXSIDE-based methods was relatively low (mean deviations of >0.5 Hz), however the tilted-EXSIDE approach provided the most accurate $^{n>1}J_{CH}$ data from a 2D-experiment in this thesis (mean deviations of <0.1 Hz for camphor).

Consequently, considering the balance between reasonable experiment time and $^{n>1}J_{CH}$ accuracy the IPAP-based methods would be recommended where the measurement of a large number of $^{n>1}J_{CH}$ values is required whereas the tilted-EXSIDE would be recommended where high accuracy of few couplings is required. However, as shown in Chapter V, to maximise the number of measurable $^{n>1}J_{CH}$ in a molecule it is necessary to combine multiple experimental methods. The combination of coupled ^{13}C, IPAP-HSQMBC and tilted-EXSIDE spectra together led to an average 40% increase in the number of measured couplings compared to using any single technique (as assessed by strychnine, camphor and 2-ethyl-1-indanone).

VII.2. Empirical prediction of $^{3}J_{CH}$ with the fragment-based approach

Chapter III and Chapter IV focussed on the relationship between DFT-calculated $^{3}J_{CH}$ and the dihedral angles Φ and Ψ. This data was used to produce a library of 556 sets of coefficients for empirically predicting $^{3}J_{CH}$ with Equation III.1 (Φ and Ψ dependant), Equation IV.5 (Φ dependant) or Equation IV.6 (Ψ dependant) that were distinguished by the coupling pathway and substituent pattern of the fragments used to generate them.

The fragments consisted of twelve different coupling pathways including saturated/unsaturated centres and β-heteroatoms; they also covered varied α, β and γ-substituents, with the primary

focus on methyl groups. The complex symmetry of the molecules formed from multiple substituents typically led to an asymmetric relationship between $^3J_{CH}$ and Φ and/or Ψ about 180° and further deviation from typical Karplus-like behaviour in terms of the exact positions (Φ and Ψ values) and relative magnitudes of the maxima and minima.

The validation of these equations by comparison to experimentally measured $^3J_{CH}$ and DFT-calculated $^3J_{CH}$ in Chapter V and Chapter VI showed several common $^3J_{CH}$ coupling pathways that would significantly improve the breadth of the fragment-based approach for organic small molecules. The focus of future extensions of the fragment-based approach should therefore be on fragments with nitrogen as a substituent or in the β-position of the coupling pathway (HCNC), vinylic HCCC coupling pathways with oxygen/nitrogen substituents and HCCC coupling pathways which include aromatic carbons (Figure VII.1). It would also be of interest to explore the effect of additional substituents such as the halogens (fluorine, chlorine and bromine), phosphorus and sulfur. However, these substituents were only present in the fragments of ~5% of the $^3J_{CH}$ in Dataset 2 and therefore are a lower priority.

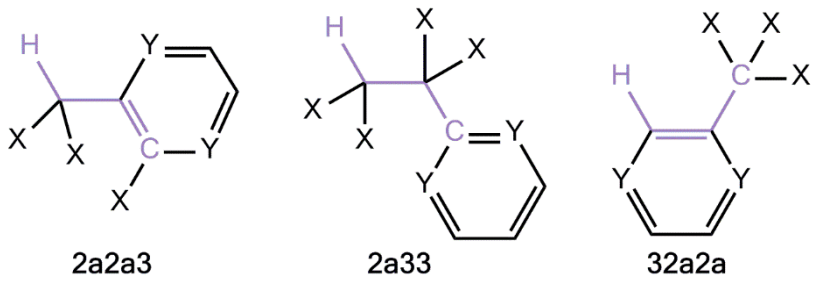


Figure VII.1 Three common aromatic $^3J_{CH}$ coupling pathways. A 6-membered ring is used to indicate the aromatic portions, however for heterocyclic compounds the ring size might necessarily vary.

VII.3. Validation of empirical $^3J_{CH}$ equations

The dataset of experimentally measured $^3J_{CH}$ described in Chapter V was used to test the performance of the fragment-based approach and selected literature empirical equations (including optimised coefficients for Aydin *et al.*). Chapter VI showed that the fragment-based approach could be improved by the inclusion of a general bond angle correction, while the optimised coefficients for Aydin *et al*. showed a similar accuracy. However, the specific bond angle correction explored for fragment HC064 showed the most accurate prediction of the DFT-calculated $^3J_{CH}$ from Dataset 1. Therefore, if larger experimental and DFT-calculated datasets of $^3J_{CH}$ were acquired, a fragment-based approach with specific bond angle corrections for each fragment may lead to further improvements in the accuracy of $^3J_{CH}$ empirical prediction. Currently, DFT-calculated $^3J_{CH}$ still offer the highest accuracy and number of $^3J_{CH}$, however empirical methods offer ~2800-fold time saving in the calculation of $^3J_{CH}$.

VII.4. Limitations and further uses of Dataset 2

Also, in Chapter VI, a second dataset (Dataset 2) was generated by searching the CSD for molecules containing a proton and carbon connected by three bonds with a carbon, nitrogen or oxygen β-atom (Figure VI.6). This resulted in 109,472 crystal structures that were then randomly sampled to give 1,002 molecules. However, as can be seen in Figure VII.2, this leads to a high population of saturated HCCC coupling pathways with proton/carbon substituents. Therefore, sampling the 109,472 structures with a 'least-similarity' approach rather than random sampling could be better suited to generating a more even coverage of different fragments that may, in turn, give a more useful assessment of the accuracy of the different empirical equations explored. For example, Palermo *et al*. (Equation V.3) predicts $^3J_{CH}$ for a wide range of β/γ-substituents (Br, NH_2, F, Cl, SH, OH), however halogen and sulfur containing fragments have a low population in Dataset 2 (<5%).

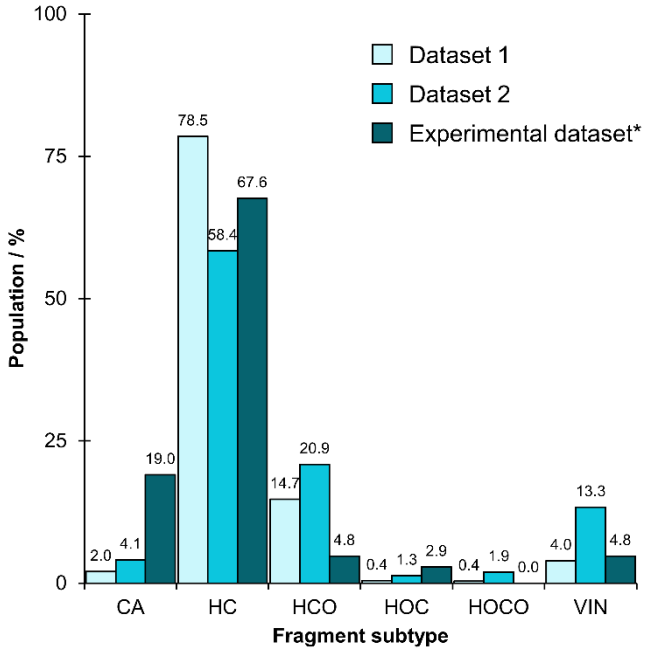


Figure VII.2 Bias in the fragments that the fragment-based approach can predict contained in the experimentally measured $^3J_{CH}$ (*strychnine, camphor and 2-ethyl-1-indanone) and the DFT-calculated $^3J_{CH}$ (Datasets 1 and 2).

The method used to estimate bond order for Dataset 2 from the Cartesian coordinates of each molecule could also be further improved by implementing a more in-depth rules-based approach, such as that described by Zhang *et al*. which achieves a 94-97% success rate in bond order estimation for small molecules[192]. However, this approach includes rules based on typical bond angle values and therefore is likely to require further adjustment to handle the variety of molecules in the CSD structures because many show a high degree of strain.

The DFT-calculations for Dataset 2 also contain chemical shifts (magnetic shielding tensors), and all $^{n}J_{XY}$ coupling constants for all pairs of atoms in each molecule. It could therefore be used to investigate the performance of, and directly generate empirical equations for other scalar couplings such as $^{3}J_{HH}$, $^{2}J_{CH}$, $^{2}J_{CC}$, $^{3}J_{CC}$, etc. The Butts group at the University of Bristol is currently exploring the use of Dataset 2 and similar DFT-calculated datasets in artificial intelligence/machine learning approaches to $^{1}H/^{13}C$ chemical shift and $^{3}J_{HH}$ prediction [182,183].

Chapter VIII. Experimental

VIII.1. NMR parameters

All spectra were recorded on a Bruker AVANCE III HD 500 MHz NMR Spectrometer with 5mm DCH ^{13}C-^{1}H/D Cryo Probe or a Varian VNMRS 500 MHz Direct Drive Spectrometer with Agilent OneNMR probe.

Spectra were processed with NMRPipe[193], and/or MestReNova 9.0.1[145] NMR processing software packages unless otherwise specified. Direct dimensions were zero-filled to 32k points and the indirect dimension twice. For the processing of IPAP spectra, a cosine-bell shaped apodization function was employed in both dimensions whereas the in-phase PIP-HSQMBC spectra subjected to lineshape fitting had exponential line broadening with 0.2 Hz in the direct dimension and cosine-bell shaped apodization function in the indirect dimension.

VIII.1.1. Sample preparation

VIII.1.1.1. Strychnine

Strychnine was purchased as (-)-strychnine 98% from Sigma Aldrich or ACROS Organics. All strychnine NMR samples were prepared as 30mg in 0.7ml of deuterated chloroform ($CDCl_3$) (130mM) in 5mm tubes under air without degassing.

VIII.1.1.2. Camphor

Camphor was purchased as camphor 96% from Aldrich. All camphor NMR samples were prepared as 30mg in 0.7ml of $CDCl_3$ (280mM) in 5mm tubes under air without degassing.

VIII.1.1.3. 2-ethyl-1-inanone

A 2% solution of 2-ethyl-1-indanone in $CDCl_3$ (120mM) in a sealed 5mm NMR tube (Agilent standard sample) was used.

VIII.1.2. Coupled ^{13}C spectra

The ^{1}H-coupled ^{13}C spectra were recorded with 3072 scans, 71680 increments, 2s relaxation delay and a spectral width of 25252.5 Hz for strychnine (total time 2h5m). The selectively ^{1}H-decoupled spectra were recorded with 1172 scans, 66560 increments, 2s relaxation delay and a spectral width of 31250.0 Hz (total time 59m) for strychnine. The ^{1}H-coupled ^{13}C spectra were recorded with 2048 scans, 65536 increments, 2s relaxation delay and a spectral width of 29761.9 Hz for camphor (total time 1h47m). The selectively ^{1}H-decoupled spectra were recorded with 1024 scans, 65536 increments, 2s relaxation delay and a spectral width of 31250.0 Hz (total time 52m) for camphor. The ^{1}H-coupled ^{13}C spectra were recorded with 512 scans, 65536 increments, 2s relaxation delay and a spectral width of 29761.9 Hz for 2-ethyl-1-indanone (total time 22m). Selective ^{1}H-decoupling of ^{13}C spectra was achieved by using an MLEV-16 supercycle[194] combined with I2Snob shaped pulses[195].

VIII.1.3. Methods that measure $^{n}J_{CH}$ from F2

Multiple coherence transfer times ($\Delta = 1 / (2 \times J_{\text{Long Range}})$) in the INEPT periods ($J_{LR}$ = 4, 6, 8 Hz) were used for the IPAP spectra and a range of J_{LR} (3-8 Hz) used for the IPAP Accordion HSQMBC. $^{1}J_{CH}$ suppression was achieved with a TANGO excitation following a GBIRD in both the forward and reverse INEPT periods for the IP-only PIP-HSQMBC[157] (fitted spectrum) and only in the forward INEPT period in the IPAP spectra. The refocused HSQMBC spectra had a GBIRD type suppression only in the forward INEPT period. The IPAP HMBC[154] spectra had a twofold low-pass filter for the same purpose. Adiabatic chirp pulses were used in the zero-quantum filter of 20-40ms length and a sweep frequency of 60 kHz that was ~9.5 times higher than the spectral width in the proton dimension.

VIII.1.3.1. PIP-HSQMBC

For strychnine, the in-phase PIP-HSQMBC spectra were recorded with 12 scans, 512 t1 increments, 12288 t2 increments, 1s relaxation delay and spectral widths of 19531.3 Hz in F1 and 6321.1 Hz in F2 (total time 3h42m for J_{LR} = 6 Hz). For camphor, the in-phase PIP-HSQMBC spectra were recorded with 8 scans, 640 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 28901.7 Hz in F1 and 4000.0 Hz in F2 (total time 3h10m for J_{LR} = 6 Hz).

VIII.1.3.2. IPAP PIP-HSQMBC

For strychnine, the IPAP PIP-HSQMBC spectra were recorded with 12 scans, 1024 t1 increments, 12288 t2 increments, 1s relaxation delay and spectral widths of 19531.3 Hz in F1 and 6321.1 Hz in F2 (total time 7h23m for J_{LR} = 6 Hz). For camphor, the IPAP PIP-HSQMBC spectra were recorded with 8 scans, 1280 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 28901.7 Hz in F1 and 4000.0 Hz in F2 (total time 6h20m for J_{LR} = 6 Hz). For 2-ethyl-1-indanone, the IPAP PIP-HSQMBC spectrum was recorded with 16 scans, 256 t1 increments, 6144 t2 increments, 1s relaxation delay and spectral widths of 26455.0 Hz in F1 and 6321.1 Hz in F2 (total time 5h3m for J_{LR} = 6 Hz).

VIII.1.3.3. IPAP refocussed HSQMBC

For strychnine, the IPAP refocussed HSQMBC spectra were recorded with 12 scans, 512 t1 increments, 12288 t2 increments, 1s relaxation delay and spectral widths of 19531.3 Hz in F1 and 6321.1 Hz in F2 (total time 3h41m for J_{LR} = 6 Hz). For camphor, the IPAP refocused HSQMBC spectra were recorded with 8 scans, 1280 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 28901.7 Hz in F1 and 4000.0 Hz in F2 (total time 6h18m for J_{LR} = 6 Hz).

VIII.1.3.4. IPAP Accordion HSQMBC

For strychnine, the IPAP Accordion HSQMBC spectrum was recorded with 4 scans, 1024 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 19531.3 Hz in F1 and 4401.4 Hz in F2 (total time 2h32m for J_{LR} = 3-8 Hz). For camphor, the IPAP Accordion HSQMBC spectrum was recorded with 4 scans, 716 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 16339.9 Hz in F1 and 4000.0 Hz in F2 (total time 2h45m for J_{LR} = 3-8 Hz).

VIII.1.3.5. IPAP HMBC

For strychnine, the IPAP HMBC spectra were recorded with 12 scans, 512 t1 increments, 12288 t2 increments, 1s relaxation delay and spectral widths of 19531.3 Hz in F1 and 6321.1 Hz in F2 (total time 3h41m for J_{LR} = 6 Hz). For camphor, the IPAP HMBC spectra were recorded with 8 scans, 1280 t1 increments, 8192 t2 increments, 1s relaxation delay and spectral widths of 28901.7 Hz in F1 and 4000.0 Hz in F2 (total time 6h28m for J_{LR} = 6 Hz).

VIII.1.4. Methods that measure $^{n}J_{CH}$ from F1

VIII.1.4.1. EXSIDE

The EXSIDE[151] spectra, which evolve $^{n}J_{CH}$ in F1, were recorded with a range of coherence transfer times (J_{LR}= 2, 4, 6, 8, 10 Hz) with the number of t1 increments chosen to give an $^{n}J_{CH}$ resolution of 1 Hz in the indirect dimension after scaling (N = 30 or 15). For strychnine, the spectra were recorded with 4 scans, 1280 t1 increments, 512 t2 increments, 1.5s relaxation delay and spectral widths of 19531.3 Hz in F1 and 2500.0 Hz in F2 for N = 30 (total time 3h44m, J_{LR} = 6 Hz). For N = 15 the spectra were recorded with 2 scans, 1592 t1 increments, 1500 t2 increments, 1.5s relaxation delay and spectral widths of 23873.5 Hz in F1 and 10000.0 Hz in F2 (total time 3h18m, J_{LR} = 8 Hz). For camphor, the spectra were recorded with 2 scans, 963 t1 increments, 1500 t2 increments, 1.5s relaxation delay and spectral widths of 28901.7 Hz in F1 and 10000 Hz in F2 for N = 30 (total time 2h15m for J_{LR} = 6 Hz). For 2-ethyl-1-indanone, the spectra were recorded with 2 scans, 1968 t1 increments, 1500 t2 increments, 1s relaxation delay and spectral widths of 29520.3 Hz in F1 and 10000.0 Hz in F2 for N = 15 (total time 4h17m for J_{LR} = 8 Hz).

VIII.1.4.2. *J*-HMBC

The *J*-HMBC[152,153] spectra were recorded with a J_{LR}= 1 Hz and the number of t1 increments chosen to give a scaling factor, κ, of 79 for strychnine and 56 for camphor. A second-order low-pass filter was used ($^{1}J_{CH}$= 120 Hz minimum to 175 Hz maximum) to supress $^{1}J_{CH}$. For strychnine the spectra were recorded with 8 scans, 256 t1 increments, 8192 t2 increments, 1.5s relaxation delay and spectral widths of 20127.5 Hz in F1 and 5498.5 Hz in F2 (total time 1h35m). For camphor the spectra were recorded with 8 scans, 512 t1 increments, 8192 t2

increments, 1.5s relaxation delay and spectral widths of 28933.0 Hz in F1 and 4000.0 Hz in F2 (total time 3h29m).

VIII.1.5. Spectral fitting procedure

Spectral fitting was used to extract accurate couplings from 1D spectra and 1D-slices (along F2) from 2D spectra. $^{n}J_{HH}$ were extracted from ^{1}H spectra and the resulting values were then used for fitting of the coupled ^{13}C spectra or slices of 2D spectra where indicated.

For camphor the spin simulation tool in MestReNova 9.0.1[145] was used to extract ^{1}H chemical shifts and $^{n}J_{HH}$ from the 1D ^{1}H spectrum. This set of data was then used to form the spin-systems necessary to extract $^{n}J_{CH}$ from each ^{13}C in the 1D coupled ^{13}C spectra. The MestReNova software does not have fitting functionality associated with the spin simulation tool and therefore the suitability of fit was assessed by eye and where the peak shape was insensitive to the $^{n}J_{CH}$ assignment the $^{n}J_{CH}$ were assigned by matching the size and sign of the experimental coupling constants to those calculated by DFT.

The analysis performed by Dr Zoltan Takacs (C4X Discovery) calculated transition frequencies and intensities from the density matrix of the spin system in question using the GAMMA software package[149]. The spectra were fitted using MINUIT2's[150] simplex algorithm[196] for regression and HESSE algorithm for error estimation. The HESSE errors reported represent a maximum range of error in $^{n}J_{CH}$ reflecting the peak linewidth and complexity. Where the fitting was found to be unstable, sensible starting parameters were identified from either a classical Lorentzian, *a priori* analysis of simpler multiplets, or with a Monte Carlo analysis[197] for complex multiplets. The fitted $^{n>1}J_{CH}$ coupling constants were also constrained within physically realistic bounds (±30Hz) to help convergence of the fitting algorithms.

VIII.2. Computational methods

The Gaussian 09 software package[78] was used for all DFT calculations, MacroModel (Version 9.9)[175] was used for conformational searching, scripts written in the AWK programming language were used to handle text data (generating inputs/processing outputs of calculations), scripts written in MATLAB R2016a[168] were used for more complex data analysis and Excel (Microsoft Office 365 ProPlus 2016) was used to collate data and perform simple statistical analysis. The Gaussian 09 and MacroModel calculations were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol[184] or the Unix computational resources of the School of Chemistry, University of Bristol.

VIII.2.1. DFT-calculation of scalar couplings for comparison to experimentally measured data.

VIII.2.1.1. Strychnine

Strychnine was geometry optimised stepwise: first with MM/UFF, then with DFT using B3LYP/3-21G, mPW1PW91/6-31G(d,p) and finally mPW1PW91/6-311G(d,p) with the IEFPCM solvent model for chloroform. NMR calculations were performed with mPW1PW91/6-311G(d,p) and the IEFPCM solvent model for chloroform using the GIAO method, including total scalar coupling constants (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms), where indicated the $^{n>1}J_{CH}$ were corrected by 6% as described in Chapter II.

The NMR calculations for strychnine were repeated (DFT, GIAO method with the IEFPCM solvent model for chloroform) with/without the “mixed” keyword and three DFT functionals (B3LYP, mPW1PW91 and ωB97X-D) combined with twelve different basis sets (3-21G,[61] 6-31G,[62-68] 6-31G(d), 6-311G(d,p)[69,70] and 6-311++G(d,p), 6-311++G(d,p)-*J*,[75,198] cc-pVDZ, cc-pVTZ, aug-cc-pVDZ and aug-cc-pVTZ[188-190] and UGBS[70,191]). The total memory allocation of the NMR calculations was 32GB over 8 processors, Gaussian 09 keywords: ‘NProcShared=8’ and ‘%mem=28GB’.

VIII.2.1.2. Camphor

Camphor was geometry optimised stepwise: first with MM/UFF, then with DFT using B3LYP/3-21G, mPW1PW91/6-31G(d,p) and finally mPW1PW91/6-311G(d,p) with the IEFPCM solvent model for chloroform. NMR calculations were performed with mPW1PW91/6-311G(d,p) and the IEFPCM solvent model for chloroform using the GIAO method, including total scalar coupling constants (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms), where indicated the $^{n>1}J_{CH}$ were corrected by 6% as described in Chapter II.

VIII.2.1.3. 2-ethyl-1-indanone

An MCMM (Monte Carlo Multiple Minimum) conformational search was performed for 2-ethyl-1-indanoe with MacroModel[175] using MMFFs (Merk Molecular Force Field)[199] with 100,000 steps, molecules within 21.0 kJ/mol of the lowest energy molecule found were retained and the structures were minimised using the TNCG (truncated Newton Conjugate Gradient) method with 500 iterations with a gradient convergence criteria of 0.05. The calculations were conducted in gaseous phase. The three conformers found (within 21.0 kJ/mol) by the conformational search were subjected to geometry optimisation and frequency calculations with DFT mPW1PW91/6-311g(d,p) and the IEFPCM solvent model for chloroform.

NMR calculations were performed with mPW1PW91/6-311G(d,p) and the IEFPCM solvent model for chloroform using the GIAO method, including total scalar coupling constants (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms); where indicated the $^{n>1}J_{CH}$ were corrected by 6% as described in Chapter II. The scalar couplings across the different conformers were then weighted according to the Boltzmann distribution (Equation V.1) using the Gibbs free energies determined for each conformer by the frequency calculations.

VIII.2.2. Generating empirical equations for $^{3}J_{CH}$

VIII.2.2.1. Varying Φ, Ψ, and/or μ

The Cartesian coordinates for the initial fragments (Φ = Ψ = 0°) were either directly drawn 'by hand' and labelled in GaussView5[177] or automatically generated by recombining the Cartesian coordinates from fully substituted fragments of the same subtype (Figure VIII.1).

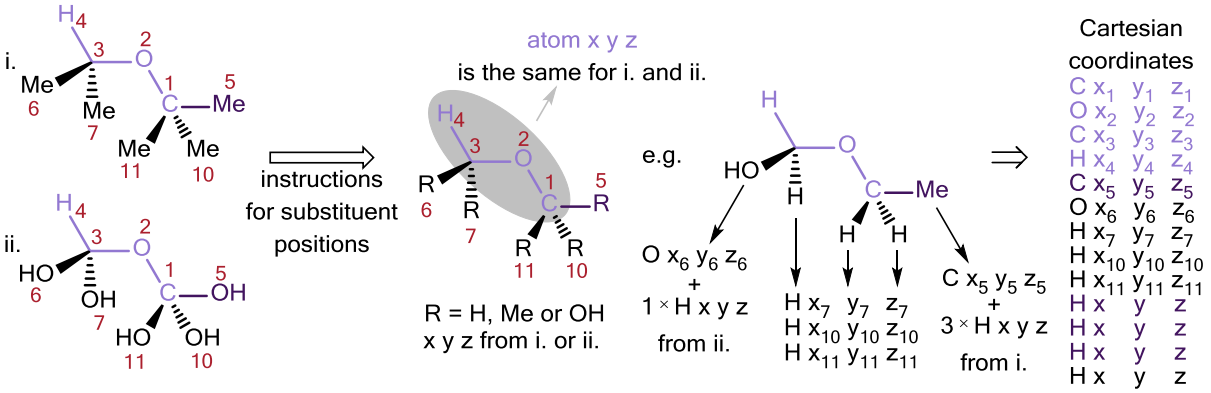


Figure VIII.1 Automation of input structure generation, using HCOC coupling pathway with H/Me/OH substituent patterns as an example, where i. and ii. show the fully substituted molecules which provide the Cartesian coordinates for recombination.

The input structures were geometry optimised by DFT with mPW1PW91/6-311g(d,p) using the 'modify redundant coordinate' functionality in Gaussian 09 to scan Φ and/or Ψ in 30° steps and freeze Φ, Ψ and/or μ. The resulting 13-2197 molecules per fragment were geometry

optimised by DFT with mPW1PW91/6-311g(d,p) using the 'modify redundant coordinate' functionality freeze Φ, Ψ and/or μ. The geometry optimised molecules were subjected to NMR calculations DFT using the GIAO method with mPW1PW91/6-311g(d,p) to determine the total scalar couplings (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms). The $^3J_{CH}$ of interest were extracted from the calculations and adjusted with the empirically determined 6% correction (Chapter II).

VIII.2.2.2. Varying Φ, Ψ, and $\theta_{1\text{-}3}$

The z-matrix for the initial fragments ($\Phi = \Psi = 0°$, $\theta_{1\text{-}3} = 120°$) were written such that Φ/Ψ (and substituent dihedral angles) and $\theta_{1\text{-}3}$ could be directly modified (in 30° and 5° steps, respectively) to generate 25,350 molecules per fragment. The input structures were geometry optimised by DFT with mPW1PW91/6-311g(d,p) using the 'modify redundant coordinate' functionality freeze Φ, Ψ and/or μ. The geometry optimised molecules were subjected to NMR calculations DFT using the GIAO method with mPW1PW91/6-311g(d,p) to determine the total scalar couplings (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms). The $^3J_{CH}$ of interest were extracted from the calculations and adjusted with the empirically determined 6% correction (Chapter II).

VIII.2.3. DFT-calculation of $^3J_{CH}$ for molecules from CSD

VIII.2.3.1. Dataset 1

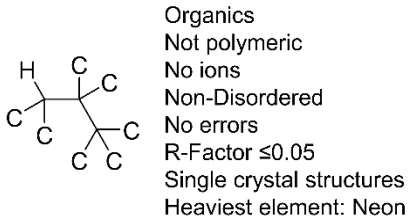


Figure VIII.2 CSD selection criteria for Dataset 1 using ConQuest 1.21[180] and CSD version 5.37[12]. Figure VI.2, repeated for clarity.

The CSD was searched for organic molecules containing three connected carbons (3° carbon - 4° carbon - 4° carbon) further substituted with carbon (Figure VI.2), which corresponds to fragment index HC064 (Appendix 4). A single molecule was selected from each of the resulting 387 structures with 3D coordinates, the molecules geometry optimised by DFT with B3LYP/6-311g(d,p), then mPW1PW91/6-311g(d,p) and subjected to NMR calculations by DFT using the GIAO method with mPW1PW91/6-311g(d,p) to determine the total scalar couplings (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms). The connectivity of the molecules was determined by using GaussView5[177]. The $^3J_{CH}$ of interest were extracted from the calculations and adjusted with the empirically determined 6% correction (Chapter II).

VIII.2.3.2. Dataset 2

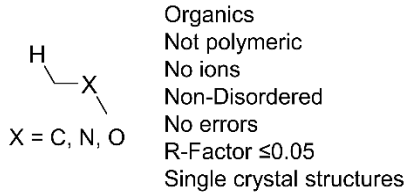


Figure VIII.3 CSD selection criteria for Dataset 2 using ConQuest 1.21[180] and CSD version 5.38[12]. Figure VI.6 repeated for clarity.

The CSD was searched for molecules containing proton and carbon atoms connected by three bonds with a carbon, nitrogen or oxygen β-atom (Figure VI.6). 1,002 structures were randomly sampled from the 109,472 crystal structures found. A single molecule was selected from each of the 1,002 structures and the molecules geometry optimised by DFT with mPW1PW91/6-311g(d,p) and subjected to NMR calculations by DFT using the GIAO method with ωB97x-D/6-311g(d,p) and the 'mixed' keyword to determine the total scalar couplings (consisting of Fermi contact, paramagnetic spin orbit, diamagnetic spin orbit and spin dipolar terms).

VIII.3. Determination of connectivity, interatomic distances, bond angle and dihedral angle with estimation of bond order

The bond angles and dihedral angles were determined for Dataset 1 and Dataset 2 (Chapter VI) from the geometry optimised Cartesian coordinates for each molecule. All interatomic distances in Dataset 2 were determined with Equation VIII.1:

$$|\boldsymbol{r_{ij}}| = \mathrm{r_{ij}} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \qquad \text{Equation VIII.1}$$

The interatomic distance (r_{ij}) between atoms i and j is the magnitude of the $\boldsymbol{r_{ij}}$ vector, where x_i, y_i and z_i are the Cartesian coordinates of atom i.

An initial estimate of connected atom pairs (i and j) was made by eliminating r_{ij} larger than the sum of the single-bond covalent radii ($r_i' + r_j'$)[200] and a scaling factor of 0.35Å[201]. The bond order (3, 2, 1.5 or 1) was estimated by assuming the following relationship between bond lengths triple bonds < double bonds < aromatic bonds < single bonds Figure VIII.4. An interatomic distance of 1.40Å was used to separate aromatic bonds between two carbon atoms from single bonds[202]. The scaling factor (0.03-0.18Å) was set by considering the mean-square deviation reported for the covalent radii (0.03Å)[200].

$r_{ij} < r_i''' + r_j''' + 0.03 \Rightarrow$ Bond order = 3

$r_{ij} < r_i'' + r_j'' + 0.06 \Rightarrow$ Bond order = 2

$r_{ij} < 1.40 + 0.03$ and i = j = carbon $\Rightarrow$ Bond order = 1.5

$r_{ij} < r_i' + r_j' + 0.18 \Rightarrow$ Bond order = 1

Figure VIII.4 Estimating bond order from interatomic distance (r_{ij}) using triple bond covalent radii (r_i''' and r_j'''), double bond covalent radii (r_i'' and r_j''), typical aromatic bond length of 1.40Å[202] and single bond covalent radii (r_i' and r_j').

The total bond order for each atom was compared to an expected value for each atom type; if they agreed the molecule was used. If pairs of connected atoms failed, the bond order was adjusted to 1.5 and the total bond order for each atom in the molecule rechecked; if the total bond order and expected total agreed the molecule was used, if the molecule failed this process it was discarded.

The bond and dihedral molecules were determined for all groups of 3-4 connected atoms in Datasets 1 and 2 as follows:

$$\theta_{ijk} = 180 - \cos^{-1}\frac{\boldsymbol{r_{ij}} \cdot \boldsymbol{r_{jk}}}{|\boldsymbol{r_{ij}}||\boldsymbol{r_{jk}}|} \qquad \text{Equation VIII.2}$$

The bond angle (θ_{ijk}) between atoms i, j and k.[203]

$$\Phi_{ijkl} = -atan2\left(\frac{\boldsymbol{r_{ij}} \cdot \boldsymbol{r_{jk}}}{|\boldsymbol{r_{ij}}||\boldsymbol{r_{jk}}|} \times \frac{\boldsymbol{r_{jk}}}{|\boldsymbol{r_{jk}}|}, \frac{\boldsymbol{r_{jk}} \times \boldsymbol{r_{kl}}}{|\boldsymbol{r_{jk}} \times \boldsymbol{r_{kl}}|}\right) \qquad \text{Equation VIII.3}$$

The dihedral angle (Φ_{ijkl}) between atoms i, j, k and l, where atan2 is a four-quadrant inverse tangent.[168,204]

VIII.4. Data analysis

VIII.4.1. Statistics

The following statistics were used to evaluate the fit between two datasets of interest, usually between one experimentally determined $^nJ_{CH}$ and one from another method such as DFT calculation. These equations use the following terms: n, the number of values in the dataset; y_i the first experimental or "true" value and x_i the second or "test" value.

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \qquad \text{Equation VIII.4}$$

Mean absolute deviation, MAD.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}}$$ Equation VIII.5

Root mean squared deviation, RMSD.

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{n}(\bar{z} - z_i)^2}{n-1}}$$ Equation VIII.6

Standard deviation, SD, where $z_i = y_i - x_i$ and $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$.

The graphs were plotted using the linear regression analysis available in Excel (Microsoft Office 365 ProPlus) which determines the gradient and intercept of a straight line with the coefficient of determination (R^2) estimating the extent of the fit.

$$\mathrm{R}^2 = \frac{[n\sum_{i=1}^{n}(x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i]^2}{[n\sum_{i=1}^{n} x_i{}^2 - (\sum_{i=1}^{n} x_i)^2][n\sum_{i=1}^{n} y_i{}^2 - (\sum_{i=1}^{n} y_i)^2]}$$ Equation VIII.7

R^2, coefficient of determination.

VIII.4.2. Regression analysis for $^3J_{CH}$ empirical equations

The coefficients in Equation III.1, Equation IV.1, Equation IV.3, Equation IV.4, Equation IV.5, Equation IV.6, Equation V.4 and Equation V.5 were determined by a least squares fit (fitlm) in MATLAB 2016[168] using DFT-calculated $^3J_{CH}$ for fragments with Φ, Ψ, μ and/or $\theta_{1\text{-}3}$ varied.

The coefficients in Equation VI.1, Equation VI.2, Equation VI.3, Equation VI.4 and Equation VI.5 were determined by minimisation of the RMSD for the DFT-calculated $^3J_{CH}$ (from Datasets 1 and 2) and empirically predicted $^3J_{CH}$ with the data solver tool available in Excel (Microsoft Office ProPlus 2016) with the GRG Nonlinear method.

Appendices.

Appendix 1. ^{1}H and ^{13}C chemical shift assignments

1.1. Strychnine

Table A.1 ^{1}H and ^{13}C chemical shifts (±0.002ppm) for strychnine in $CDCl_3$ at 25°C. The chemical shifts are referenced relative to internal chloroform at 7.284ppm (^{1}H) and 77.070ppm (^{13}C).[125]

^{1}H	*pro-R* / *pro-S*	δ /ppm
H13		1.280
H15a	*pro-R*	1.460
H17b	*pro-R*	1.893
H17a	*pro-S*	1.897
H15b	*pro-S*	2.364
H11b	*pro-R*	2.679
H20b	*pro-S*	2.729
H18b	*pro-R*	2.875
H11a	*pro-S*	3.141
H14		3.148
H18a	*pro-S*	3.196
H20a	*pro-S*	3.708
H8		3.865
H16		3.944
H23a	*pro-S*	4.071
H23b	*pro-R*	4.156
H12		4.293
H22		5.903
H2		7.102
H1		7.168
H3		7.261
H4		8.102

^{13}C	δ /ppm
C15	26.878
C14	31.619
C11	42.508
C17	42.898
C13	48.247
C18	50.363
C7	51.957
C20	52.689
C8	60.127
C16	60.208
C23	64.625
C12	77.618
C4	116.229
C1	122.281
C2	124.212
C22	127.191
C3	128.548
C6	132.812
C21	140.615
C5	142.218
C10	169.312

1.2. Camphor

Table A.2 ^{1}H and ^{13}C chemical shifts (±0.002ppm) for camphor in $CDCl_3$ at 25°C. The chemical shifts are referenced relative to internal chloroform at 7.284ppm (^{1}H) and 77.044ppm (^{13}C).[125]

^{1}H	*pro-R* / *pro-S**	δ /ppm
H10		0.834
H6		0.913
H9		0.960
H3b	*pro-S*	1.341
H4b	*pro-S*	1.405
H4a	*pro-R*	1.683
H1b	*pro-S*	1.845
H3a	*pro-R*	1.952
H2		2.090
H1a	*pro-R*	2.351

^{13}C	δ /ppm
C6	9.249
C9	19.144
C10	19.779
C3	27.048
C4	29.914
C2	43.045
C1	43.299
C8	46.789
C5	57.697
C7	219.684

* Defined for the enantiomer in Figure II.2

1.3. 2-ethyl-1-indanone

Table A.3 ^{1}H and ^{13}C chemical shifts (±0.002ppm) for 2-ethyl-1-indanone in $CDCl_3$ at 25°C. The chemical shifts are referenced relative to internal chloroform at 7.2589ppm (^{1}H) and 76.992ppm (^{13}C).

^{1}H	***pro-R* / *pro-S****	**δ /ppm**
H9		1.009
H8a	*pro-R*	1.540
H8b	*pro-S*	1.972
H7		2.614
H6b	*pro-S*	2.820
H6a	*pro-R*	3.316
H2		7.356
H4		7.457
H3		7.574
H1		7.745

* Defined for the enantiomer in Figure V.1

^{13}C	**δ /ppm**
C9	11.587
C8	24.457
C6	32.322
C7	48.745
C1	123.812
C4	126.517
C2	127.270
C3	134.586
C11	136.942
C5	153.807
C10	208.935

Appendix 2. Experimentally measured $^{n}J_{CH}$

2.1. Strychnine

Table A.4 Summary of all calculated/experimental/literature $^{n}J_{CH}$ values for which a value could be measured in coupled ^{13}C spectra. (a) Absolute DFT + 6% correction (b) Coupled ^{13}C (c) IPAP accordion HSQMBC, J_{LR}=3-8Hz (d) IPAP PIP-HSQMBC, J_{LR}=6Hz (e) IPAP analysis of refocused HSQMBC, J_{LR}=6Hz (f) IPAP HSQMBC-COSY, J_{LR}=6Hz (g) IPAP HMBC, J_{LR}=6Hz (h) IPAP HMBC-COSY, J_{LR}=6Hz (i) EXSIDE, J_{LR}=6Hz, N=30, tilted with equation Equation II.1 (j) *J*-HMBC, J_{LR}=1Hz, κ=79 (k) Lineshape analysis of PIP-HSQMBC, J_{LR}=8Hz (l) Literature[129,133,134,137,138,166].[125]

$^{n}J_{CH}$ /Hz		(a) DFT	(b) Coupled ^{13}C (Error†)	(c) IPAP ac. HSQMBC	(d) IPAP PIP-HSQMBC	(e) IPAP Refocussed HSQMBC	(f) IPAP HSQMBC-COSY	(g) IPAP HMBC	(h) IPAP HMBC-COSY	(i) EXSIDE (tilted)	(j) *J*-HMBC	(k) PIP-HSQMBC	(l) Literature (SD)
C1	H2	0.7	0.9 (0.6)	1.4				1.4			0.9		0.8
C1	H3	8.5	8.9 (0.4)	8.8	9.1			9.0			9.2		8.9 (0.4)
C1	H4	1.4	1.1 (1.1)	1.3		0.6	0.6	1.8	1.6				1.2
C1	H8	0.3	1.3 (1.3)	1.3	1.6		0.9						
C2	H1	0.5	0.3 (0.9)										0.9
C2	H3	0.4	1.3 (0.9)								1.8		1.9
C2	H4	7.4	7.6 (0.4)					7.6		7.6	7.8		7.6 (0.3)
C4	H1	1.4	1.1 (0.2)			1.5	1.7	1.8	1.7		1.1		1.0
C4	H2	7.6	8.0 (0.3)	7.7	7.9	7.7		7.8			7.9		7.9
C4	H3	1.2	2.1 (0.2)	2.3	2.3			2.2	2.4		2.0		2.4 (0.1)
C4	H8	0.1	0.4 (0.5)										
C5	H1	7.9	8.0 (0.9)	8.0	8.0	8.3	8.2	7.9			8.1		
C5	H2	1.6	1.3 (0.5)	1.9		1.4	1.3	1.6	1.5		0.8		
C5	H3	9.7	9.7 (1.0)	10.0	10.1			9.8			10.0		
C5	H4	3.4	2.2 (1.1)	2.2			2.3	2.9		2.2	2.0		1.1 (0.4)
C5	H8	3.0	3.2 (1.0)	3.2	3.2	3.3	3.2	3.3	3.3	3.1	3.0	3.4	3.1 (0.4)
C5	H11b	0.7	0.6 (0.6)				0.1	1.3			0.8		
C5	H12	0.1	1.0 (1.0)						0.4				
C5	H13	0.1	0.4 (0.6)			0.3	0.4		0.1				
C10	H8	0.1	0.6 (0.7)										
C10	H11a	6.6	6.4 (0.5)	6.4	6.4	6.4	6.4		6.4	6.4	6.5	6.5	6.3 (0.3)
C10	H11b	8.4	7.9 (0.5)	8.0	7.9	7.9	8.0	7.9	8.0	7.9	8.0		7.6 (0.4)
C10	H12	1.5	2.0 (0.5)	2.1	2.1	1.9	2.1	2.6	2.1		1.8	2.4	2.9 (1.9)

C11	H12	2.0	1.9 (0.3)	2.2	1.8						1.7		2.1
C11	H13	0.7	0.6 (1.1)										0.6
C15	H13	3.8	3.2 (1.8)	3.4	3.5			3.6		3.4	3.3		4.4 (1.3)
C15	H14	3.9	4.0 (1.7)			1.9	2.6		4.9	3.8			3.1
C15	H16	2.0	1.2 (1.5)	2.3				2.6					2.2
C15	H20b	1.4	1.9 (0.9)	2.0	1.7	1.8	1.9		1.7		1.4		
C17	H8	6.3	5.9 (0.7)	5.9	5.9	5.9			6.0	5.8	5.7		5.7 (0.3)
C17	H13	0.1	0.7 (1.3)			0.2	0.1		0.1				0.6
C17	H18a	1.7	1.6 (0.7)	2.6					2.6				3.7 (0.9)
C17	H18b	4.7	4.6 (0.5)	5.1	4.7	4.7	4.7			4.8	5.0		5.0 (0.5)
C17	H20b	0.1	1.0 (1.0)										
C22	H14	3.8	3.5 (1.6)							3.9			
C22	H15a	0.5	0.9 (0.9)										
C22	H20a	4.8	5.4 (0.6)	5.5	5.4	5.5	5.4	5.4	5.8	5.4	5.3	5.5	5.3 (0.6)
C22	H20b	4.8	4.7 (1.1)	4.7	4.6	4.6	4.7	4.8	4.7	4.7	4.7	4.7	4.8 (0.5)
C22	H23a	3.4	3.2 (1.1)	3.4	3.2	3.2		3.3	3.3		3.4		3.9 (0.3)
C22	H23b	4.1	4.0 (1.1)	3.8		3.7	3.4	3.8	3.7		3.7		3.6 (0.2)
C23	H12	3.1	2.7 (0.4)	2.6	2.6	3.0	2.7	2.9		2.7	2.8		3.5 (2.2)
C23	H20a	1.5	1.2 (1.2)	3.0	2.3						1.1		
C23	H20b	0.3	0.9 (1.6)										0.2
C23	H22	6.4	7.0 (0.6)	7.2				6.9	7.5	7.0		7.3	6.4 (0.5)
$^{n}J_{CH}$ Total		44	44	29	20	20	20	23	20	14	27	6	29
$^{n}J_{CH}$ <1 Hz		14	11	0	0	3	5	0	3	0	3	0	5

† HESSE error in fitting of $^{n}J_{CH}$ reported from fitting fully coupled ^{13}C spectrum.

2.1. Camphor

Table A.5 Comparison of different coupled ^{13}C spectrum analyses. (a) MestReNova spin simulation including δ(^{1}H) and $^{n}J_{HH}$ and simulations of selectively decoupled peaks. (b) MestReNova spin simulation including δ(^{1}H) and $^{n}J_{HH}$ performed by Alexander Jeffreys[147]. (c) Dr Zoltan Takacs, C4X Discovery simulation[148]. (d) Zoltan Takacs, C4X Discovery simulation, spin simulation of the methyl ^{1}H decoupled spectrum[148].

			DFT	**(a) MestReNova - fully coupled and methyl decoupled ^{13}C spectra**	**(b) MestReNova - fully coupled ^{13}C spectra**	**(c) C4XD - methyl decoupled ^{13}C spectra**	**(d) C4XD - Fully coupled ^{13}C spectra**
^{1}H	**^{13}C**	**n**	**$^{n}J_{HC}$ / Hz**	**$^{n}J_{HC}$ / Hz**	**$^{n}J_{HC}$ / Hz**	**$^{n}J_{HC}$ (Error†) / Hz**	**$^{n}J_{HC}$ (Error†) / Hz**
H6	C6	1	115.41	126.0	125.9		126.0 (0.3)
H4a	C6	3	3.01	3.0	3.0		3.0 (0.2)
H2	C6	4	0.46				0.6 (0.2)
H10	C9	3	4.56	5.0	5.0		4.9 (0.3)
H9	C9	1	114.58	125.0	124.6		125.0 (2.2)
H1b	C9	4	0.83	1.0	1.0		1.0 (0.5)
H2	C9	3	1.21	1.2	1.2		1.2 (0.7)
H10	C10	1	114.51	124.8	125.0		124.9 (1.3)
H9	C10	3	4.57	4.9	4.7		5.0 (1.1)
H1b	C10	4	0.35				0.6 (0.6)
H2	C10	3	0.86				0.9 (0.4)
H3b	C3	1	123.08		134.3		
H3a	C3	1	121.31		130.8		
H6	C4	4	4.03	4.0	4.0		4.3 (0.5)
H3b	C4	2	-2.05	3.0	3.2	2.2 (0.4)	2.3 (0.3)
H4b	C4	1	123.40	133.3	132.5	133.1 (0.2)	134.3 (0.4)
H4a	C4	1	124.20	133.7	134.1	129.6 (0.2)	132.7 (0.4)
H3a	C4	2	-3.70	3.6	3.8	4.0 (0.4)	4.0 (0.4)
H2	C4	3	6.26	7.5	7.5	6.8 (0.3)	6.9 (0.3)
H3b	C2	2	-2.87			0.3 (2.0)	
H4b	C2	3	1.50			3.2 (0.8)	
H4a	C2	3	1.26			1.6 (0.6)	
H1b	C2	2	-2.25			2.9 (0.8)	
H3a	C2	2	-3.02			2.0 (0.7)	
H2	C2	1	129.90	142.5	142.8	142.4 (0.7)	
H1a	C2	2	-2.29			2.7 (0.6)	

H3b	C1	3	5.27	5.1	5.0	5.1 (0.2)	5.2 (0.2)
H4b	C1	4	-0.20			0.6 (0.4)	0.7 (0.3)
H4a	C1	4	-0.77			0.3 (0.9)	0.4 (0.4)
H1b	C1	1	123.60	133.2	133.4	132.3 (0.2)	133.1 (0.1)
H3a	C1	3	8.46	9.1	9.2	9.2 (0.2)	9.3 (0.2)
H2	C1	2	-0.75			0.4 (0.5)	0.5 (0.3)
H1a	C1	1	120.63	130.0	130.0	129.7 (0.2)	130.0 (0.1)
H3b	C8	3	5.96	6.1		6.1 (0.7)	
H4b	C8	3	5.71	6.1		6.5 (0.6)	
H1b	C8	3	8.08	8.5		8.5 (0.4)	
H1a	C8	3	0.57			0.6 (0.9)	
H6	C7	3	3.67				4.2 (1.5)
H3b	C7	4	-0.32			1.0 (0.7)	0.8 (2.0)
H4b	C7	3	4.81			4.6 (1.6)	4.1 (1.6)
H4a	C7	3	6.88			8.1 (0.6)	8.2 (0.9)
H1b	C7	2	-4.65			4.6 (1.5)	4.3 (1.0)
H3a	C7	4	-0.84			0.1 (1.9)	0.3 (1.4)
H2	C7	3	8.02			8.4 (0.6)	8.5 (1.3)
H1a	C7	2	-6.40			6.4 (0.3)	6.4 (0.2)

† HESSE error in fitting of $^{n}J_{CH}$ reported from fitting fully coupled or selectively decoupled ^{13}C spectrum.

Table A.6 Summary of all calculated/experimental/literature $^{n}J_{CH}$ values for which a value could be measured in coupled ^{13}C spectra. (a) Absolute DFT + 6% correction (b) Coupled ^{13}C (fully or selectively coupled) (c) IPAP accordion HSQMBC, J_{LR}=3-8Hz (d) IPAP PIP-HSQMBC, J_{LR}=6Hz (e) IPAP analysis of refocused HSQMBC, J_{LR}=6Hz (g) IPAP HMBC, J_{LR}=6Hz (i) EXSIDE, J_{LR}=6Hz, N=30, tilted with Equation II.1 (j) *J*-HMBC, J_{LR}=1Hz, κ=56.[125]

$^{n}J_{CH}$ /Hz		(a) DFT	(b) Coupled ^{13}C (Error†)	(c) IPAP accordion HSQMBC	(d) IPAP PIP-HSQMBC	(e) IPAP Refocussed HSQMBC	(g) IPAP HMBC	(i) EXSIDE (tilted)	(j) *J*-HMBC
C1	H2	0.8	0.5 (0.3)						
C1	H3a	9.0	9.3 (0.2)						9.1
C1	H3b	5.6	5.2 (0.2)						
C1	H4a	0.8	0.4 (0.4)						
C1	H4b	0.2	0.7 (0.3)						
C2	H1a	2.4	2.7 (**0.6**)					2.4	2.5
C2	H1b	2.4	2.9 (**0.8**)					2.7	2.4
C2	H3a	3.2	2.0 (**0.7**)						2.8
C2	H3b	3.0	0.3 (**2.0**)‡						
C2	H4a	1.3	1.6 (**0.6**)						
C2	H4b	1.6	3.2 (**0.8**)						1.1
C4	H2	6.6	6.9 (0.3)	6.8	6.8	6.9	6.8	7.0	
C4	H3a	3.9	4.0 (0.4)			3.6	3.9	3.6	3.6
C4	H3b	2.2	2.3 (0.3)	3.1	3.1	3.3	3.2		
C4	H6	4.3	4.3 (0.5)	4.4	4.4	4.4	4.4	4.5	4.5
C6	H2	0.5	0.6 (0.2)	1.9	1.6	1.5	1.8		
C6	H4a	3.2	3.0 (0.2)	3.1	2.9	3.0	3.2	3.1	2.7
C7	H1a	6.8	6.4 (0.2)	6.4	6.3	6.4	6.4	6.5	6.4
C7	H1b	4.9	4.3 (1.0)	4.7	4.6	4.7	4.7	4.7	4.6
C7	H2	8.5	8.5 (1.3)	8.4	8.6	8.7	8.4	8.6	1.0 §
C7	H3a	0.9	0.3 (1.4)				0.9		1.0
C7	H3b	0.3	0.8 (2.0)‡				0.2		
C7	H4a	7.3	8.2 (0.9)	7.7	7.8	7.6	8.0	7.8	7.4
C7	H4b	5.1	4.1 (1.6)	4.5		4.5	4.6		4.4
C7	H6	3.9	4.2 (1.5)	4.0	4.0	4.0	4.1	4.1	4.0
C8	H1a	0.6	0.6 (**0.9**)	1.1	0.6	1.6	0.8		1.4
C8	H1b	8.6	8.5 (**0.4**)	8.6	8.6	8.5	8.5	8.6	8.8
C8	H3b	6.3	6.1 (**0.7**)	6.5	6.4	6.5	6.6		6.4
C8	H4b	6.0	6.5 (**0.6**)	6.3	6.4	6.3	6.3		6.4

C9	H10	4.8	4.9 (0.3)	4.7	4.7	4.8	4.5	5.0	4.9
C9	H1b	0.9	1.0 (0.5)	1.4			1.6		1.0
C9	H2	1.3	1.2 (0.7)	2.3	1.7	1.9	1.9		1.5
C10	H1b	0.4	0.6 (0.6)	1.3			1.4		
C10	H2	0.9	0.9 (0.4)	1.9	1.6	1.7	1.7		
C10	H9	4.8	5.0 (1.1)	4.9	4.9	4.9	4.9	5.1	5.0
$^{n}J_{CH}$ Total		35	35	21	18	20	24	14	23
$^{n}J_{CH}$<1Hz		10	10	0	1	0	3	0	0

† HESSE error in fitting of $^{n}J_{CH}$ reported from fitting fully coupled or selectively decoupled (**bold**) ^{13}C spectrum. ‡ Not included in MAD/SD calculation due to large (≥2.0Hz) HESSE error. § MAD/SD calculated including this value are 0.42/1.14Hz

2.2. 2-ethyl-1-indanone

Table A.7 Summary of all calculated/experimental $^{n}J_{CH}$ values (a) absolute $^{n>1}J_{CH}$ calculated by DFT + 6% correction (b) Coupled ^{13}C (fully coupled) (c) IPAP PIP-HSQMBC, J_{LR}=6Hz (d) EXSIDE, J_{LR}=8Hz, N=15, (e) tilted with Equation II.1.

$^{n}J_{CH}$ /Hz		(a) DFT	(b) Coupled ^{13}C	(c) IPAP PIP-HSQMBC	(d) EXSIDE	(e) Tilted EXSIDE
C1	H1		162.3			
C1	H2	0.38	1.2	1.8		
C1	H3	7.75	7.9	7.6	8.3	8.3
C1	H4	1.50	1.2	2.3		
C1	H6a	1.04	1.2	2.0		
C1	H6b	1.50	1.2	1.9		
C1	H7	0.07				
C1	H8a	0.01				
C1	H8b	0.03				
C1	H9	0.04				
C2	H1	0.23	1.2			
C2	H2		160.3			
C2	H3	0.42	1.2			
C2	H4	6.90	7.3		7.1	7.3
C2	H6a	1.03	1.3			
C2	H6b	1.40	1.3			
C2	H7	0.21				
C2	H8a	0.05				
C2	H8b	0.03				
C2	H9	0.04				
C3	H1	7.35	7.7		7.7	7.8
C3	H2	0.98	1.3			
C3	H3		158.3			
C3	H4	0.34	1.3			
C3	H6a	1.48	1.6			
C3	H6b	1.97	1.7			
C3	H7	0.12				
C3	H8a	0.03				
C3	H8b	0.02				
C3	H9	0.03				
C4	H1	1.54	1.5			
C4	H2	7.41	7.9		7.7	7.8

C4	H3	0.12				
C4	H4		159.4			
C4	H6a	2.41	2.2		1.9	2.9
C4	H6b	2.13	2.1		1.2	2.3
C4	H7	0.29				
C4	H8a	0.12				
C4	H8b	0.27				
C4	H9	0.02				
C5	H1	5.95	6.4		5.9	6.3
C5	H2	1.48	1.2			
C5	H3	7.99	6.4	7.9	8.2	8.2
C5	H4	0.02				
C5	H6a	5.59	6.4	5.8	5.3	5.8
C5	H6b	7.00	7.0	6.3	6.2	6.5
C5	H7	1.03	2.2	2.1	1.3	2.4
C5	H8a	0.09				
C5	H8b	0.12				
C5	H9	0.04				
C6	H1	0.07				
C6	H2	0.55				
C6	H3	0.47				
C6	H4	3.13	3.0	3.8	2.2	3.2
C6	H6a		132.1			
C6	H6b		129.5			
C6	H7	3.99	3.2			
C6	H8a	4.48	5.3		4.3	5.0
C6	H8b	6.30	5.6	5.9	5.7	6.1
C6	H9	0.11				
C7	H1	0.16				
C7	H2	0.19				
C7	H3	0.14				
C7	H4	0.22				
C7	H6a	2.82	3.6		2.2	3.2
C7	H6b	3.77	3.6			
C7	H7		127.9			
C7	H8a	4.39	4.9	4.6	3.6	4.4
C7	H8b	4.45	4.9		3.6	4.3
C7	H9	4.89	4.8	5.1	4.6	5.2
C8	H1	0.03				

C8	H2	0.03				
C8	H3	0.02				
C8	H4	0.01				
C8	H6a	0.70	1.8	2.3		
C8	H6b	4.99	5.2	5.1	4.5	5.1
C8	H7	5.93	5.3		5.4	5.8
C8	H8a		126.9			
C8	H8b		128.2			
C8	H9	4.26	4.2	4.4	3.7	4.4
C9	H1	0.03				
C9	H2	0.00				
C9	H3	0.02				
C9	H4	0.02				
C9	H6a	0.17				
C9	H6b	0.13				
C9	H7	4.83	5.0	5.1	4.6	5.2
C9	H8a	3.96	4.2		3.5	4.3
C9	H8b	4.14	4.2		3.7	4.4
C9	H9		124.9			
C10	H1	2.78		1.9	1.8	2.8
C10	H2	0.62				
C10	H3	0.70				
C10	H4	0.20				
C10	H6a	4.63		3.6	2.9	3.7
C10	H6b	0.42		1.9		
C10	H7	7.36		6.1	5.9	6.2
C10	H8a	2.25			1.4	2.5
C10	H8b	3.52		4.0	3.3	4.1
C10	H9	0.35				
C11	H1	0.07				
C11	H2	7.41	6.9	7.5	7.7	7.8
C11	H3	1.39	1.0	2.1		
C11	H4	5.89	6.9	6.5	6.0	6.3
C11	H6a	2.70	2.8	2.8	2.0	3.0
C11	H6b	2.25	2.8	2.6	1.6	2.7
C11	H7	0.63	1.0			
C11	H8a	0.07				
C11	H8b	0.01				
C11	H9	0.19				

Appendix 3. Computational models of molecules

3.1. Strychnine

Table A.8 Cartesian coordinates for the major conformer of strychnine,[174] geometry optimised using DFT with mPW1PW91/6-311g(d,p) IEFPCM(Chloroform)[125] as described in Chapter II.

Atom	Coordinates /Å		
	x	y	z
C	2.3427	0.4386	0.1691
C	3.6511	0.8757	-0.0033
C	4.6113	-0.0720	-0.3503
C	4.2799	-1.4126	-0.5125
C	2.9655	-1.8358	-0.3203
C	1.9971	-0.9064	0.0196
C	0.5472	-1.1175	0.3628
C	-0.0027	0.3346	0.4100
C	-0.2662	-2.0669	-0.5489
C	-1.0942	-2.2352	1.6990
C	0.3666	-1.8152	1.7234
C	-0.8170	-1.4029	-1.8018
C	-1.7295	-0.2475	-1.3817
C	-0.8095	0.8267	-0.7920
C	-2.7004	-2.2408	-0.1466
C	-2.7990	-0.7671	-0.4359
C	-3.6937	0.0324	0.1431
C	-3.6870	1.5222	-0.0381
C	-1.4109	2.1847	-0.4155
C	-0.3576	3.0513	0.3057
C	1.0707	2.5453	0.3122
N	1.1978	1.1921	0.4768
N	-1.3556	-2.6009	0.3032
O	2.0240	3.2974	0.2044
O	-2.5063	2.1223	0.4807
H	3.8992	1.9183	0.1203
H	5.6363	0.2487	-0.4952
H	5.0445	-2.1309	-0.7822
H	2.7088	-2.8832	-0.4323
H	-0.6020	0.4765	1.3133
H	0.3942	-2.8925	-0.8389
H	-1.7343	-1.3986	2.0171
H	-1.2965	-3.0748	2.3699
H	1.0208	-2.6902	1.7629
H	0.6168	-1.1676	2.5655
H	0.0085	-1.0341	-2.4158
H	-1.3541	-2.1363	-2.4069
H	-2.2202	0.1771	-2.2663
H	-0.0829	1.0488	-1.5847
H	-2.9495	-2.8265	-1.0377
H	-3.4138	-2.5346	0.6253
H	-4.4322	-0.3914	0.8178
H	-4.5144	1.9768	0.5073
H	-3.8023	1.7945	-1.0974
H	-1.7398	2.6861	-1.3357
H	-0.6810	3.1207	1.3492
H	-0.3285	4.0661	-0.0858

3.2. Camphor

Table A.9 Cartesian coordinates for camphor, geometry optimised using DFT with mPW1PW91/6-311g (d,p) IEFPCM(Chloroform)[125] as described in Chapter II.

Atom	Coordinates /Å		
	x	y	z
C	-0.8591	1.2469	1.0648
C	0.6007	1.2570	0.5965
C	0.5890	1.7702	-0.8525
C	-0.0673	0.5981	-1.6238
C	-0.2913	-0.5016	-0.5442
C	-0.5424	-1.8875	-1.0841
C	-1.4336	0.0711	0.2826
C	0.9108	-0.2563	0.4200
C	2.2755	-0.5319	-0.2063
C	0.8193	-1.0678	1.7139
H	0.0212	2.6984	-0.9418
H	1.5968	1.9748	-1.2164
H	-1.0050	0.8777	-2.1089
H	0.5867	0.2045	-2.4045
H	-1.4059	2.1649	0.8373
H	-0.9737	1.0607	2.1350
H	1.2810	1.7909	1.2614
H	1.0077	-2.1251	1.5119
H	1.5765	-0.7312	2.4271
H	-0.1527	-1.0026	2.2060
H	3.0681	-0.2013	0.4710
H	2.4120	-1.6038	-0.3698
H	2.4274	-0.0302	-1.1618
H	-1.4600	-1.9032	-1.6762
H	0.2805	-2.2161	-1.7229
H	-0.6637	-2.6141	-0.2776
O	-2.5758	-0.3168	0.3036

3.3. 2-ethyl-1-indanone

Table A.10 Cartesian coordinates for 2-ethyl-1-indanone, geometry optimised using DFT with mPW1PW91/6-311g(d,p) IEFPCM(Chloroform) as described in Chapter V.

Atom	Conformer 1 Coordinates /Å			Conformer 2 Coordinates /Å			Conformer 3 Coordinates /Å		
	x	y	z	x	y	z	x	y	z
C	0.9409	-0.8621	-0.0647	-0.7495	-0.8304	0.0728	0.7650	-0.8356	-0.2195
C	0.7386	0.5147	-0.0464	-0.8442	0.5585	0.0498	0.6683	0.5508	-0.1333
C	1.7970	1.4145	0.0493	-2.0692	1.2121	-0.0585	1.7691	1.3495	0.1671
C	3.0842	0.9073	0.1145	-3.2185	0.4428	-0.1325	2.9910	0.7316	0.3736
C	3.2962	-0.4753	0.0885	-3.1321	-0.9530	-0.1022	3.0977	-0.6603	0.2819
C	2.2343	-1.3682	0.0015	-1.9057	-1.5994	-0.0019	1.9941	-1.4523	-0.0122
H	1.6040	2.4808	0.0655	-2.1083	2.2950	-0.0776	1.6586	2.4257	0.2302
H	3.9324	1.5776	0.1830	-4.1890	0.9175	-0.2118	3.8695	1.3219	0.6044
H	4.3097	-0.8563	0.1385	-4.0410	-1.5409	-0.1589	4.0615	-1.1291	0.4446
H	2.4171	-2.4366	-0.0122	-1.8576	-2.6822	0.0158	2.0954	-2.5297	-0.0759
C	-0.3669	-1.6019	-0.1275	0.6841	-1.2829	0.1549	-0.5687	-1.4630	-0.5254
H	-0.3700	-2.3666	-0.9072	0.8317	-2.0213	0.9457	-0.5081	-2.1399	-1.3805
H	-0.5479	-2.1178	0.8215	0.9771	-1.7638	-0.7845	-0.9043	-2.0638	0.3250
C	-1.4337	-0.5154	-0.3629	1.4928	0.0076	0.3809	-1.5144	-0.2765	-0.7877
H	-1.7075	-0.5146	-1.4263	1.7635	0.0774	1.4430	-1.6611	-0.1818	-1.8708
C	-2.7095	-0.6887	0.4644	2.7640	0.1595	-0.4483	-2.8967	-0.3310	-0.1355
H	-2.4754	-0.5011	1.5183	3.1285	1.1821	-0.3200	-3.4334	-1.2020	-0.5235
H	-3.0076	-1.7404	0.4023	2.5096	0.0520	-1.5087	-3.4523	0.5503	-0.4671
C	-3.8702	0.1920	0.0204	3.8533	-0.8357	-0.0728	-2.8794	-0.3658	1.3884
H	-4.1420	-0.0193	-1.0178	3.5246	-1.8681	-0.2170	-2.3980	0.5250	1.7985
H	-4.7523	0.0108	0.6387	4.7467	-0.6867	-0.6828	-3.8974	-0.4001	1.7816
H	-3.6107	1.2478	0.0900	4.1427	-0.7227	0.9756	-2.3523	-1.2410	1.7761
C	-0.7034	0.8110	-0.1385	0.5022	1.1493	0.1514	-0.7196	0.9700	-0.3969
O	-1.2119	1.9085	-0.0607	0.7885	2.3239	0.0774	-1.1646	2.0950	-0.3307

Table A.11 Populations of 2-ethyl-1-indanone conformers; frequency calculations to determine the relative energies used DFT with mPW1PW91/6-311g (d,p) IEFPCM(Chloroform), as described in Chapter V.

Conformer number	Relative Energy (DFT) /kJ.mol^{-1}	Population
1	4.547	11%
2	0.000	69%
3	3.150	19%

Appendix 4. $^3J_{CH}$ empirical equations

4.1. Labelling systems

The structures are defined in Table A.12 to Table A.19 using a structural identifier containing atom types and bond orders (BO), where the vertical pipes "|" separate the coupling pathway and substituents:

coupling pathway | α-substituents | β-substituents | γ-substituents

The coupling pathway up to the first pipe is defined as:

α [space] α-β BO [space] β [space] β-γ BO [space] γ [space]

The order of the substituents in the subsequent sections indicates their orientation relative to the coupling pathway, as defined for each subtype in Figure A.1. The α_1 substituent is used to define the dihedral angle Ψ (γ-β-α-α_1).

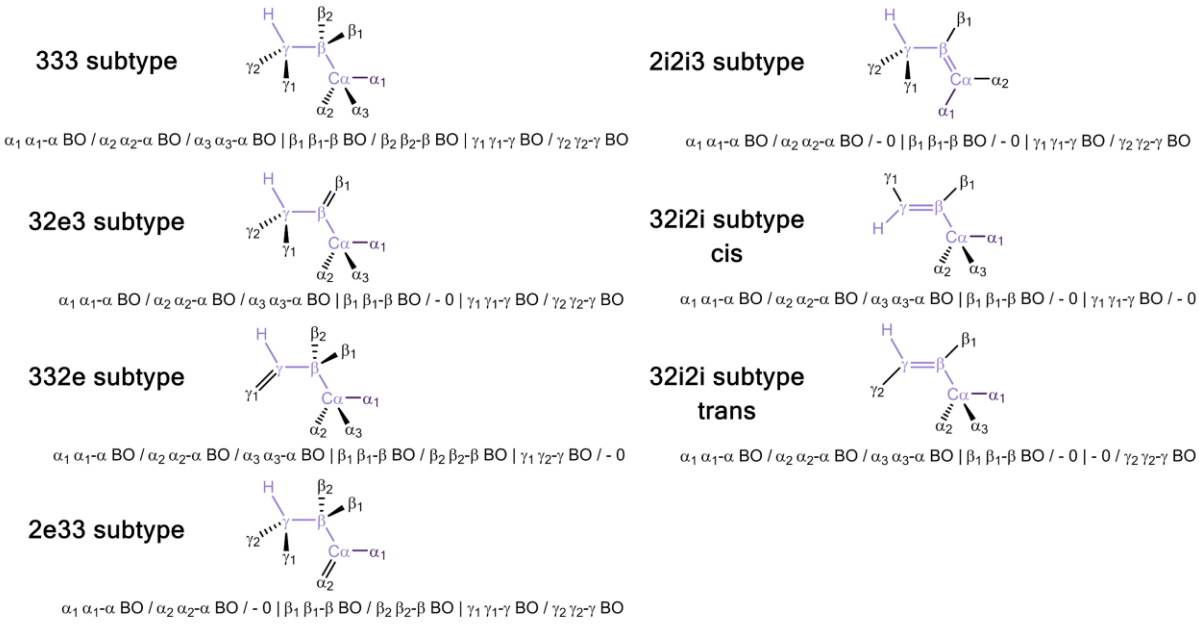


Figure A.1 Substituent orientation definitions in the structural identifiers in Table A.12 to Table A.19.

In Table A.12 to Table A.19 the substituents in the fragments used to generate the empirical equations were as follows: H = proton, C = methyl group, CA = aldehyde, O = primary alcohol, - = no substituent (sp^2), lp = lone pair (for O/N β position).

Table A.12 Definition of structures for HCCC saturated hydrocarbon fragments. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HC001	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | H 1 / H 1
HC002	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | H 1 / H 1
HC003	333	C 1 C 1 C | C 1 / H 1 / C 1 | H 1 / H 1 | H 1 / H 1
HC004	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / H 1
HC005	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | H 1 / H 1
HC006	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | H 1 / H 1
HC007	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1
HC008	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 1 / H 1
HC009	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | H 1 / C 1
HC010	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1
HC011	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / H 1
HC012	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / H 1
HC013	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1
HC014	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | H 1 / H 1
HC015	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | H 1 / H 1
HC016	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1
HC017	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | H 1 / H 1
HC018	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | H 1 / H 1
HC019	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1
HC020	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | H 1 / C 1
HC021	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 1 / H 1
HC022	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1
HC023	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1
HC024	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1
HC025	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 1 / C 1
HC026	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1
HC027	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1
HC028	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 1 / C 1
HC029	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1
HC030	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1
HC031	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | C 1 / C 1
HC032	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1
HC033	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1
HC034	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | C 1 / C 1
HC035	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / C 1
HC036	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / H 1
HC037	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1
HC038	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1
HC039	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1
HC040	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / C 1
HC041	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | H 1 / C 1
HC042	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | C 1 / H 1
HC043	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | C 1 / C 1
HC044	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | H 1 / C 1
HC045	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 1 / H 1
HC046	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 1 / C 1
HC047	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1
HC048	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1
HC049	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / C 1
HC050	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1
HC051	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1
HC052	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / C 1
HC053	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1
HC054	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1
HC055	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 1 / C 1
HC056	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | H 1 / C 1
HC057	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / H 1
HC058	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1
HC059	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / C 1
HC060	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | C 1 / H 1
HC061	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | C 1 / C 1
HC062	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / C 1

HC063	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / H 1
HC064	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / C 1
HC065	333	C 1 C 1 C | C 1 / C 1 / H 1 | H 1 / H 1 | H 1 / H 1
HC066	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / H 1

Table A.13 Definition of structures for HCCC fragments with mixed H/C/CA substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HCA001	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / H 1 | H 1 / H 1
HCA002	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / H 1 | H 1 / H 1
HCA003	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / H 1 | H 1 / H 1
HCA004	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / H 1 | H 1 / H 1
HCA005	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / CA 1 | H 1 / H 1
HCA006	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / H 1 | H 1 / H 1
HCA007	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / C 1 | H 1 / H 1
HCA008	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / CA 1 | H 1 / H 1
HCA009	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | CA 1 / H 1
HCA010	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | H 1 / CA 1
HCA011	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | CA 1 / C 1
HCA012	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 1 / CA 1
HCA013	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / H 1 | H 1 / H 1
HCA014	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / C 1 | H 1 / H 1
HCA015	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1
HCA016	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / H 1 | H 1 / H 1
HCA017	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / C 1 | H 1 / H 1
HCA018	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / C 1 | H 1 / H 1
HCA019	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / H 1 | H 1 / H 1
HCA020	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / C 1 | H 1 / H 1
HCA021	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / C 1 | H 1 / H 1
HCA022	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / H 1 | H 1 / H 1
HCA023	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / C 1 | H 1 / H 1
HCA024	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1
HCA025	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / H 1 | H 1 / H 1
HCA026	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / CA 1 | H 1 / H 1
HCA027	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / CA 1 | H 1 / H 1
HCA028	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / C 1 | H 1 / H 1
HCA029	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / H 1 | H 1 / H 1
HCA030	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / CA 1 | H 1 / H 1
HCA031	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / CA 1 | H 1 / H 1
HCA032	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / C 1 | H 1 / H 1
HCA033	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / H 1 | H 1 / H 1
HCA034	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / CA 1 | H 1 / H 1
HCA035	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / C 1 | H 1 / H 1
HCA036	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / CA 1 | H 1 / H 1
HCA037	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / H 1 | H 1 / C 1
HCA038	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / H 1 | C 1 / H 1
HCA039	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1
HCA040	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / H 1 | H 1 / C 1
HCA041	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / H 1 | C 1 / H 1
HCA042	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / H 1 | C 1 / C 1
HCA043	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / H 1 | H 1 / C 1
HCA044	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / H 1 | C 1 / H 1
HCA045	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / H 1 | C 1 / C 1
HCA046	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1
HCA047	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1
HCA048	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / H 1 | C 1 / C 1
HCA049	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | H 1 / CA 1
HCA050	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / CA 1
HCA051	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / CA 1
HCA052	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | CA 1 / H 1
HCA053	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | CA 1 / H 1
HCA054	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | CA 1 / H 1
HCA055	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 1 / CA 1

HCA056	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 1 / CA 1
HCA057	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 1 / CA 1
HCA058	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | CA 1 / C 1
HCA059	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | CA 1 / C 1
HCA060	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | CA 1 / C 1
HCA061	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / CA 1 | H 1 / C 1
HCA062	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / CA 1 | C 1 / H 1
HCA063	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / CA 1 | C 1 / C 1
HCA064	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / H 1 | H 1 / C 1
HCA065	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / H 1 | C 1 / H 1
HCA066	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / H 1 | C 1 / C 1
HCA067	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / CA 1 | H 1 / C 1
HCA068	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / CA 1 | C 1 / H 1
HCA069	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / CA 1 | C 1 / C 1
HCA070	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / C 1 | H 1 / C 1
HCA071	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / C 1 | C 1 / H 1
HCA072	333	C 1 C 1 C | H 1 / H 1 / H 1 | CA 1 / C 1 | C 1 / C 1
HCA073	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | CA 1 / H 1
HCA074	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | CA 1 / H 1
HCA075	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | CA 1 / H 1
HCA076	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | H 1 / CA 1
HCA077	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | H 1 / CA 1
HCA078	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / CA 1
HCA079	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | CA 1 / C 1
HCA080	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | CA 1 / C 1
HCA081	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | CA 1 / C 1
HCA082	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | C 1 / CA 1
HCA083	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | C 1 / CA 1
HCA084	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / CA 1
HCA085	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1
HCA086	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1
HCA087	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / C 1 | H 1 / C 1
HCA088	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / H 1 | H 1 / C 1
HCA089	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / C 1 | H 1 / C 1
HCA090	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / C 1 | H 1 / C 1
HCA091	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / H 1 | H 1 / C 1
HCA092	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / C 1 | H 1 / C 1
HCA093	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / C 1 | H 1 / C 1
HCA094	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1
HCA095	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / C 1 | H 1 / C 1
HCA096	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / C 1 | H 1 / C 1
HCA097	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1
HCA098	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1
HCA099	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / C 1 | C 1 / H 1
HCA100	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / H 1 | C 1 / H 1
HCA101	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / C 1 | C 1 / H 1
HCA102	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / C 1 | C 1 / H 1
HCA103	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / H 1 | C 1 / H 1
HCA104	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / C 1 | C 1 / H 1
HCA105	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / C 1 | C 1 / H 1
HCA106	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1
HCA107	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / C 1 | C 1 / H 1
HCA108	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / C 1 | C 1 / H 1
HCA109	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / H 1 | C 1 / C 1
HCA110	333	C 1 C 1 C | CA 1 / H 1 / H 1 | H 1 / C 1 | C 1 / C 1
HCA111	333	C 1 C 1 C | CA 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1
HCA112	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / H 1 | C 1 / C 1
HCA113	333	C 1 C 1 C | H 1 / C 1 / CA 1 | H 1 / C 1 | C 1 / C 1
HCA114	333	C 1 C 1 C | H 1 / C 1 / CA 1 | C 1 / C 1 | C 1 / C 1
HCA115	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / H 1 | C 1 / C 1
HCA116	333	C 1 C 1 C | H 1 / CA 1 / C 1 | H 1 / C 1 | C 1 / C 1
HCA117	333	C 1 C 1 C | H 1 / CA 1 / C 1 | C 1 / C 1 | C 1 / C 1
HCA118	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / H 1 | C 1 / C 1
HCA119	333	C 1 C 1 C | CA 1 / C 1 / C 1 | H 1 / C 1 | C 1 / C 1
HCA120	333	C 1 C 1 C | CA 1 / C 1 / C 1 | C 1 / C 1 | C 1 / C 1
HCA121	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / H 1 | H 1 / C 1
HCA122	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / CA 1 | H 1 / C 1

HCA123	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / CA 1 | H 1 / C 1
HCA124	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / H 1 | H 1 / C 1
HCA125	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / CA 1 | H 1 / C 1
HCA126	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / CA 1 | H 1 / C 1
HCA127	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / H 1 | H 1 / C 1
HCA128	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / CA 1 | H 1 / C 1
HCA129	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / CA 1 | H 1 / C 1
HCA130	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / C 1 | H 1 / C 1
HCA131	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / C 1 | H 1 / C 1
HCA132	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / C 1 | H 1 / C 1
HCA133	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / H 1 | C 1 / H 1
HCA134	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / CA 1 | C 1 / H 1
HCA135	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / CA 1 | C 1 / H 1
HCA136	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / H 1 | C 1 / H 1
HCA137	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / CA 1 | C 1 / H 1
HCA138	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / CA 1 | C 1 / H 1
HCA139	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / H 1 | C 1 / H 1
HCA140	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / CA 1 | C 1 / H 1
HCA141	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / CA 1 | C 1 / H 1
HCA142	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / C 1 | C 1 / H 1
HCA143	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / C 1 | C 1 / H 1
HCA144	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / C 1 | C 1 / H 1
HCA145	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / H 1 | C 1 / C 1
HCA146	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / CA 1 | C 1 / C 1
HCA147	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / CA 1 | C 1 / C 1
HCA148	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / H 1 | C 1 / C 1
HCA149	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / CA 1 | C 1 / C 1
HCA150	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / CA 1 | C 1 / C 1
HCA151	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / H 1 | C 1 / C 1
HCA152	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / CA 1 | C 1 / C 1
HCA153	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / CA 1 | C 1 / C 1
HCA154	333	C 1 C 1 C | H 1 / C 1 / C 1 | CA 1 / C 1 | C 1 / C 1
HCA155	333	C 1 C 1 C | C 1 / H 1 / H 1 | CA 1 / C 1 | C 1 / C 1
HCA156	333	C 1 C 1 C | C 1 / C 1 / C 1 | CA 1 / C 1 | C 1 / C 1
HCA157	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / CA 1
HCA158	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | H 1 / CA 1
HCA159	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | H 1 / CA 1
HCA160	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | CA 1 / H 1
HCA161	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | CA 1 / H 1
HCA162	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | CA 1 / H 1
HCA163	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / CA 1
HCA164	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | C 1 / CA 1
HCA165	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 1 / CA 1
HCA166	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | CA 1 / C 1
HCA167	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | CA 1 / C 1
HCA168	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | CA 1 / C 1
HCA169	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / CA 1
HCA170	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | H 1 / CA 1
HCA171	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | H 1 / CA 1
HCA172	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | CA 1 / H 1
HCA173	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | CA 1 / H 1
HCA174	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | CA 1 / H 1
HCA175	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / CA 1
HCA176	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / CA 1
HCA177	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 1 / CA 1
HCA178	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | CA 1 / C 1
HCA179	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | CA 1 / C 1
HCA180	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | CA 1 / C 1
HCA181	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | H 1 / CA 1
HCA182	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / CA 1
HCA183	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / CA 1
HCA184	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | CA 1 / H 1
HCA185	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | CA 1 / H 1
HCA186	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | CA 1 / H 1
HCA187	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / CA 1
HCA188	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | C 1 / CA 1
HCA189	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / CA 1

HCA190	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | CA 1 / C 1
HCA191	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | CA 1 / C 1
HCA192	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | CA 1 / C 1

Table A.14 Definition of structures for HCCC fragments with mixed H/C/O substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HCO001	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / H 1 | H 1 / H 1
HCO002	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / H 1 | H 1 / H 1
HCO003	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / H 1 | H 1 / H 1
HCO004	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / H 1 | H 1 / H 1
HCO005	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / O 1 | H 1 / H 1
HCO006	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / H 1 | H 1 / H 1
HCO007	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / C 1 | H 1 / H 1
HCO008	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / O 1 | H 1 / H 1
HCO009	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | O 1 / H 1
HCO010	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | H 1 / O 1
HCO011	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | O 1 / C 1
HCO012	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 1 / O 1
HCO013	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / H 1 | H 1 / H 1
HCO014	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / C 1 | H 1 / H 1
HCO015	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1
HCO016	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / H 1 | H 1 / H 1
HCO017	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / C 1 | H 1 / H 1
HCO018	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / C 1 | H 1 / H 1
HCO019	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / H 1 | H 1 / H 1
HCO020	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / C 1 | H 1 / H 1
HCO021	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / C 1 | H 1 / H 1
HCO022	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / H 1 | H 1 / H 1
HCO023	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / C 1 | H 1 / H 1
HCO024	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1
HCO025	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / H 1 | H 1 / H 1
HCO026	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / O 1 | H 1 / H 1
HCO027	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / O 1 | H 1 / H 1
HCO028	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / C 1 | H 1 / H 1
HCO029	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / H 1 | H 1 / H 1
HCO030	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | H 1 / H 1
HCO031	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / O 1 | H 1 / H 1
HCO032	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / C 1 | H 1 / H 1
HCO033	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / H 1 | H 1 / H 1
HCO034	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / O 1 | H 1 / H 1
HCO035	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / C 1 | H 1 / H 1
HCO036	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / O 1 | H 1 / H 1
HCO037	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / H 1 | H 1 / C 1
HCO038	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / H 1 | C 1 / H 1
HCO039	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1
HCO040	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / H 1 | H 1 / C 1
HCO041	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / H 1 | C 1 / H 1
HCO042	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / H 1 | C 1 / C 1
HCO043	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / H 1 | H 1 / C 1
HCO044	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / H 1 | C 1 / H 1
HCO045	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / H 1 | C 1 / C 1
HCO046	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1
HCO047	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1
HCO048	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / H 1 | C 1 / C 1
HCO049	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | H 1 / O 1
HCO050	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / O 1
HCO051	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / O 1
HCO052	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | O 1 / H 1
HCO053	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | O 1 / H 1
HCO054	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | O 1 / H 1
HCO055	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 1 / O 1
HCO056	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 1 / O 1

HCO057	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 1 / O 1
HCO058	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | O 1 / C 1
HCO059	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | O 1 / C 1
HCO060	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | O 1 / C 1
HCO061	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / O 1 | H 1 / C 1
HCO062	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / O 1 | C 1 / H 1
HCO063	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / O 1 | C 1 / C 1
HCO064	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / H 1 | H 1 / C 1
HCO065	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / H 1 | C 1 / H 1
HCO066	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / H 1 | C 1 / C 1
HCO067	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / O 1 | H 1 / C 1
HCO068	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / O 1 | C 1 / H 1
HCO069	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / O 1 | C 1 / C 1
HCO070	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / C 1 | H 1 / C 1
HCO071	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / C 1 | C 1 / H 1
HCO072	333	C 1 C 1 C | H 1 / H 1 / H 1 | O 1 / C 1 | C 1 / C 1
HCO073	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | O 1 / H 1
HCO074	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | O 1 / H 1
HCO075	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | O 1 / H 1
HCO076	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | H 1 / O 1
HCO077	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | H 1 / O 1
HCO078	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / O 1
HCO079	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | O 1 / C 1
HCO080	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | O 1 / C 1
HCO081	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | O 1 / C 1
HCO082	333	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | C 1 / O 1
HCO083	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | C 1 / O 1
HCO084	333	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / O 1
HCO085	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1
HCO086	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1
HCO087	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / C 1 | H 1 / C 1
HCO088	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / H 1 | H 1 / C 1
HCO089	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / C 1 | H 1 / C 1
HCO090	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / C 1 | H 1 / C 1
HCO091	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / H 1 | H 1 / C 1
HCO092	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / C 1 | H 1 / C 1
HCO093	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / C 1 | H 1 / C 1
HCO094	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1
HCO095	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / C 1 | H 1 / C 1
HCO096	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / C 1 | H 1 / C 1
HCO097	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1
HCO098	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1
HCO099	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / C 1 | C 1 / H 1
HCO100	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / H 1 | C 1 / H 1
HCO101	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / C 1 | C 1 / H 1
HCO102	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / C 1 | C 1 / H 1
HCO103	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / H 1 | C 1 / H 1
HCO104	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / C 1 | C 1 / H 1
HCO105	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / C 1 | C 1 / H 1
HCO106	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1
HCO107	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / C 1 | C 1 / H 1
HCO108	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / C 1 | C 1 / H 1
HCO109	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / H 1 | C 1 / C 1
HCO110	333	C 1 C 1 C | O 1 / H 1 / H 1 | H 1 / C 1 | C 1 / C 1
HCO111	333	C 1 C 1 C | O 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1
HCO112	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / H 1 | C 1 / C 1
HCO113	333	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / C 1 | C 1 / C 1
HCO114	333	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / C 1 | C 1 / C 1
HCO115	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / H 1 | C 1 / C 1
HCO116	333	C 1 C 1 C | H 1 / O 1 / C 1 | H 1 / C 1 | C 1 / C 1
HCO117	333	C 1 C 1 C | H 1 / O 1 / C 1 | C 1 / C 1 | C 1 / C 1
HCO118	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / H 1 | C 1 / C 1
HCO119	333	C 1 C 1 C | O 1 / C 1 / C 1 | H 1 / C 1 | C 1 / C 1
HCO120	333	C 1 C 1 C | O 1 / C 1 / C 1 | C 1 / C 1 | C 1 / C 1
HCO121	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / H 1 | H 1 / C 1
HCO122	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / O 1 | H 1 / C 1
HCO123	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / O 1 | H 1 / C 1

HCO124	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / H 1 | H 1 / C 1
HCO125	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | H 1 / C 1
HCO126	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / O 1 | H 1 / C 1
HCO127	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / H 1 | H 1 / C 1
HCO128	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / O 1 | H 1 / C 1
HCO129	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / O 1 | H 1 / C 1
HCO130	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / C 1 | H 1 / C 1
HCO131	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / C 1 | H 1 / C 1
HCO132	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / C 1 | H 1 / C 1
HCO133	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / H 1 | C 1 / H 1
HCO134	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | C 1 / H 1
HCO135	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / O 1 | C 1 / H 1
HCO136	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / H 1 | C 1 / H 1
HCO137	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / O 1 | C 1 / H 1
HCO138	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / O 1 | C 1 / H 1
HCO139	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / H 1 | C 1 / H 1
HCO140	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / O 1 | C 1 / H 1
HCO141	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / O 1 | C 1 / H 1
HCO142	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / C 1 | C 1 / H 1
HCO143	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / C 1 | C 1 / H 1
HCO144	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / C 1 | C 1 / H 1
HCO145	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / H 1 | C 1 / C 1
HCO146	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / O 1 | C 1 / C 1
HCO147	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / O 1 | C 1 / C 1
HCO148	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / H 1 | C 1 / C 1
HCO149	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / O 1 | C 1 / C 1
HCO150	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / O 1 | C 1 / C 1
HCO151	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / H 1 | C 1 / C 1
HCO152	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | C 1 / C 1
HCO153	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / O 1 | C 1 / C 1
HCO154	333	C 1 C 1 C | C 1 / C 1 / C 1 | O 1 / C 1 | C 1 / C 1
HCO155	333	C 1 C 1 C | H 1 / C 1 / C 1 | O 1 / C 1 | C 1 / C 1
HCO156	333	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / C 1 | C 1 / C 1
HCO157	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / O 1
HCO158	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | H 1 / O 1
HCO159	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | H 1 / O 1
HCO160	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | O 1 / H 1
HCO161	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | O 1 / H 1
HCO162	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | O 1 / H 1
HCO163	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / O 1
HCO164	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | C 1 / O 1
HCO165	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 1 / O 1
HCO166	333	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | O 1 / C 1
HCO167	333	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | O 1 / C 1
HCO168	333	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | O 1 / C 1
HCO169	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / O 1
HCO170	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | H 1 / O 1
HCO171	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | H 1 / O 1
HCO172	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | O 1 / H 1
HCO173	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | O 1 / H 1
HCO174	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | O 1 / H 1
HCO175	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / O 1
HCO176	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / O 1
HCO177	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 1 / O 1
HCO178	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | O 1 / C 1
HCO179	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | O 1 / C 1
HCO180	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | O 1 / C 1
HCO181	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | H 1 / O 1
HCO182	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / O 1
HCO183	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / O 1
HCO184	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | O 1 / H 1
HCO185	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | O 1 / H 1
HCO186	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | O 1 / H 1
HCO187	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / O 1
HCO188	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | C 1 / O 1
HCO189	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / O 1
HCO190	333	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | O 1 / C 1

HCO191	333	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | O 1 / C 1
HCO192	333	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | O 1 / C 1

Table A.15 Definition of structures for carbonyl containing HCCC fragments with mixed H/C/O substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
CA001	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | O 2 / - 0 | H 1 / H 1
CA002	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | O 2 / - 0 | H 1 / H 1
CA003	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | O 2 / - 0 | H 1 / H 1
CA004	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | H 1 / H 1
CA005	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | O 2 / - 0 | C 1 / H 1
CA006	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | O 2 / - 0 | H 1 / C 1
CA007	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | O 2 / - 0 | C 1 / C 1
CA008	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | O 2 / - 0 | C 1 / H 1
CA009	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | O 2 / - 0 | H 1 / C 1
CA010	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | O 2 / - 0 | C 1 / C 1
CA011	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | O 2 / - 0 | C 1 / H 1
CA012	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | O 2 / - 0 | H 1 / C 1
CA013	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | O 2 / - 0 | C 1 / C 1
CA014	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | C 1 / H 1
CA015	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | H 1 / C 1
CA016	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | C 1 / C 1
CA017	332e	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | O 2 / - 0
CA018	332e	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | O 2 / - 0
CA019	332e	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | O 2 / - 0
CA020	332e	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | O 2 / - 0
CA021	332e	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | O 2 / - 0
CA022	332e	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | O 2 / - 0
CA023	332e	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | O 2 / - 0
CA024	332e	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | O 2 / - 0
CA025	332e	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | O 2 / - 0
CA026	332e	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | O 2 / - 0
CA027	332e	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | O 2 / - 0
CA028	332e	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | O 2 / - 0
CA029	332e	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | O 2 / - 0
CA030	332e	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | O 2 / - 0
CA031	332e	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | O 2 / - 0
CA032	332e	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | O 2 / - 0
CA033	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / H 1 | H 1 / H 1
CA034	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / H 1 | H 1 / H 1
CA035	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / H 1 | H 1 / H 1
CA036	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / C 1 | H 1 / H 1
CA037	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / C 1 | H 1 / H 1
CA038	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / H 1 | H 1 / C 1
CA039	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / H 1 | C 1 / H 1
CA040	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / H 1 | C 1 / C 1
CA041	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | H 1 / H 1
CA042	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | H 1 / H 1
CA043	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | H 1 / H 1
CA044	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / H 1 | H 1 / C 1
CA045	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / H 1 | C 1 / H 1
CA046	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / H 1 | C 1 / C 1
CA047	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / C 1 | H 1 / C 1
CA048	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / H 1 | H 1 / C 1
CA049	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / C 1 | H 1 / C 1
CA050	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / C 1 | C 1 / H 1
CA051	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / H 1 | C 1 / H 1
CA052	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / C 1 | C 1 / H 1
CA053	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | H 1 / C 1 | C 1 / C 1
CA054	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / H 1 | C 1 / C 1
CA055	2e33	C 1 C 1 C | H 1 / O 2 / - 0 | C 1 / C 1 | C 1 / C 1
CA056	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | H 1 / C 1
CA057	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | H 1 / C 1

CA058	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | H 1 / C 1
CA059	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | C 1 / H 1
CA060	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | C 1 / H 1
CA061	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | C 1 / H 1
CA062	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | C 1 / C 1
CA063	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | C 1 / C 1
CA064	2e33	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | C 1 / C 1

Table A.16 Definition of structures for HCOC fragments with mixed H/C substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HOC001	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / H 1
HOC002	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / H 1
HOC005	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / H 1
HOC006	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / H 1
HOC007	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / H 1
HOC008	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / C 1
HOC009	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / C 1
HOC010	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / H 1
HOC011	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / C 1
HOC012	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / C 1
HOC013	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / H 1
HOC014	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / C 1
HOC015	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / C 1
HOC016	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / H 1
HOC017	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / C 1
HOC018	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / C 1

Table A.17 Definition of structures for HCNC fragments with mixed H/C substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HNC001	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / H 1 | H 1 / H 1
HNC002	333	C 1 N 1 C | H 1 / H 1 / H 1 | H 1 / LP 0 | H 1 / H 1
HNC003	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / H 1 | H 1 / H 1
HNC004	333	C 1 N 1 C | C 1 / H 1 / H 1 | H 1 / LP 0 | H 1 / H 1
HNC005	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / H 1 | H 1 / H 1
HNC006	333	C 1 N 1 C | H 1 / C 1 / C 1 | H 1 / LP 0 | H 1 / H 1
HNC007	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / H 1 | H 1 / H 1
HNC008	333	C 1 N 1 C | C 1 / C 1 / C 1 | H 1 / LP 0 | H 1 / H 1
HNC013	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / H 1 | C 1 / H 1
HNC014	333	C 1 N 1 C | H 1 / H 1 / H 1 | H 1 / LP 0 | C 1 / H 1
HNC015	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / H 1 | H 1 / C 1
HNC016	333	C 1 N 1 C | H 1 / H 1 / H 1 | H 1 / LP 0 | H 1 / C 1
HNC017	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / H 1 | C 1 / C 1
HNC018	333	C 1 N 1 C | H 1 / H 1 / H 1 | H 1 / LP 0 | C 1 / C 1
HNC019	333	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / H 1
HNC020	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / C 1 | H 1 / H 1
HNC021	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / C 1 | H 1 / H 1
HNC022	333	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / H 1
HNC023	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / C 1 | H 1 / H 1
HNC024	333	C 1 N 1 C | C 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / H 1
HNC025	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / H 1 | H 1 / C 1
HNC026	333	C 1 N 1 C | C 1 / H 1 / H 1 | H 1 / LP 0 | H 1 / C 1
HNC027	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / H 1 | C 1 / H 1
HNC028	333	C 1 N 1 C | C 1 / H 1 / H 1 | H 1 / LP 0 | C 1 / H 1
HNC029	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / H 1 | C 1 / C 1
HNC030	333	C 1 N 1 C | C 1 / H 1 / H 1 | H 1 / LP 0 | C 1 / C 1
HNC031	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / H 1 | H 1 / C 1
HNC032	333	C 1 N 1 C | H 1 / C 1 / C 1 | H 1 / LP 0 | H 1 / C 1
HNC033	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / H 1 | C 1 / H 1

HNC034	333	C 1 N 1 C | H 1 / C 1 / C 1 | H 1 / LP 0 | C 1 / H 1
HNC035	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / H 1 | C 1 / C 1
HNC036	333	C 1 N 1 C | H 1 / C 1 / C 1 | H 1 / LP 0 | C 1 / C 1
HNC037	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / H 1 | H 1 / C 1
HNC038	333	C 1 N 1 C | C 1 / C 1 / C 1 | H 1 / LP 0 | H 1 / C 1
HNC039	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / H 1 | C 1 / H 1
HNC040	333	C 1 N 1 C | C 1 / C 1 / C 1 | H 1 / LP 0 | C 1 / H 1
HNC041	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / H 1 | C 1 / C 1
HNC042	333	C 1 N 1 C | C 1 / C 1 / C 1 | H 1 / LP 0 | C 1 / C 1
HNC043	333	C 1 N 1 C | H 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / C 1
HNC044	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / C 1 | C 1 / H 1
HNC045	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / C 1 | C 1 / C 1
HNC046	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / C 1 | H 1 / C 1
HNC047	333	C 1 N 1 C | H 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / H 1
HNC048	333	C 1 N 1 C | H 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / C 1
HNC049	333	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / C 1
HNC050	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / C 1 | H 1 / C 1
HNC051	333	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / H 1
HNC052	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / C 1 | C 1 / H 1
HNC053	333	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / C 1
HNC054	333	C 1 N 1 C | C 1 / H 1 / H 1 | LP 0 / C 1 | C 1 / C 1
HNC055	333	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / C 1
HNC056	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / C 1 | H 1 / C 1
HNC057	333	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / H 1
HNC058	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / C 1 | C 1 / H 1
HNC059	333	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / C 1
HNC060	333	C 1 N 1 C | H 1 / C 1 / C 1 | LP 0 / C 1 | C 1 / C 1
HNC061	333	C 1 N 1 C | C 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / C 1
HNC062	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / C 1 | H 1 / C 1
HNC063	333	C 1 N 1 C | C 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / H 1
HNC064	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / C 1 | C 1 / H 1
HNC065	333	C 1 N 1 C | C 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / C 1
HNC066	333	C 1 N 1 C | C 1 / C 1 / C 1 | LP 0 / C 1 | C 1 / C 1
HNC067	333	C 1 N 1 C | H 1 / H 1 / H 1 | LP 0 / C 1 | H 1 / H 1
HNC068	333	C 1 N 1 C | H 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / H 1

Table A.18 Definition of structures for HCOC fragments with mixed H/C/O substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
HOCO001	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / H 1
HOCO002	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | H 1 / H 1
HOCO003	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | H 1 / H 1
HOCO004	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / H 1
HOCO005	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / H 1
HOCO006	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / O 1
HOCO007	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / C 1
HOCO008	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / O 1
HOCO009	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / C 1
HOCO010	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / H 1
HOCO011	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / C 1
HOCO012	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | H 1 / C 1
HOCO013	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | C 1 / H 1
HOCO014	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | C 1 / C 1
HOCO015	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | H 1 / C 1
HOCO016	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | C 1 / H 1
HOCO017	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | C 1 / C 1
HOCO018	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / C 1
HOCO019	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / H 1
HOCO020	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / C 1
HOCO021	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / O 1
HOCO022	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / O 1
HOCO023	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / O 1
HOCO024	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / H 1

HOCO025	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / H 1
HOCO026	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / H 1
HOCO027	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / O 1
HOCO028	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / O 1
HOCO029	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / O 1
HOCO030	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / C 1
HOCO031	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / C 1
HOCO032	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / C 1
HOCO033	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / H 1
HOCO034	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / H 1
HOCO035	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / C 1
HOCO036	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / C 1
HOCO037	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / H 1
HOCO038	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / H 1
HOCO039	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / C 1
HOCO040	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / C 1
HOCO041	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / H 1
HOCO042	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | H 1 / O 1
HOCO043	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / C 1
HOCO044	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / O 1
HOCO045	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | O 1 / H 1
HOCO046	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | H 1 / O 1
HOCO047	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | O 1 / C 1
HOCO048	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | C 1 / O 1
HOCO049	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | O 1 / H 1
HOCO050	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | H 1 / O 1
HOCO051	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | O 1 / C 1
HOCO052	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | C 1 / O 1
HOCO053	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / H 1
HOCO054	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / O 1
HOCO055	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / C 1
HOCO056	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / O 1
HOCO057	333	C 1 O 1 C | H 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / O 1
HOCO058	333	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / O 1
HOCO059	333	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / O 1
HOCO060	333	C 1 O 1 C | C 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / O 1
HOCO061	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / H 1
HOCO062	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / H 1
HOCO063	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / C 1
HOCO064	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / C 1
HOCO065	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / H 1
HOCO066	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / O 1
HOCO067	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / C 1
HOCO068	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / O 1
HOCO069	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / H 1
HOCO070	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / O 1
HOCO071	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / C 1
HOCO072	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / O 1
HOCO073	333	C 1 O 1 C | O 1 / H 1 / H 1 | LP 0 / LP 0 | O 1 / O 1
HOCO074	333	C 1 O 1 C | H 1 / O 1 / C 1 | LP 0 / LP 0 | O 1 / O 1
HOCO075	333	C 1 O 1 C | H 1 / C 1 / O 1 | LP 0 / LP 0 | O 1 / O 1
HOCO076	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | H 1 / O 1
HOCO077	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / H 1
HOCO078	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | C 1 / O 1
HOCO079	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / C 1
HOCO080	333	C 1 O 1 C | H 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / O 1
HOCO081	333	C 1 O 1 C | C 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / O 1
HOCO082	333	C 1 O 1 C | O 1 / C 1 / C 1 | LP 0 / LP 0 | O 1 / O 1
HOCO083	333	C 1 O 1 C | O 1 / O 1 / O 1 | LP 0 / LP 0 | O 1 / O 1

Table A.19 Definition of structures for vinyl containing HCCC fragments with mixed H/C substituents. Subtype as defined in Figure III.3, structural identifier as defined in Figure A.1.

Fragment Index	Subtype	Structural identifier
VIN001	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / H 1 | H 1 / H 1
VIN002	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / H 1 | H 1 / H 1
VIN003	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / H 1 | H 1 / H 1
VIN004	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / H 1 | H 1 / H 1
VIN005	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / C 1 | H 1 / H 1
VIN006	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | H 1 / H 1
VIN007	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / C 1 | H 1 / H 1
VIN008	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / C 1 | H 1 / H 1
VIN009	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / H 1 | C 1 / H 1
VIN010	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / H 1 | C 1 / H 1
VIN011	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / H 1 | C 1 / H 1
VIN012	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / H 1 | C 1 / H 1
VIN013	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / C 1 | C 1 / H 1
VIN014	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | C 1 / H 1
VIN015	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / C 1 | C 1 / H 1
VIN016	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / C 1 | C 1 / H 1
VIN017	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / H 1 | H 1 / C 1
VIN018	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / H 1 | H 1 / C 1
VIN019	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / H 1 | H 1 / C 1
VIN020	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / H 1 | H 1 / C 1
VIN021	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / C 1 | H 1 / C 1
VIN022	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | H 1 / C 1
VIN023	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / C 1 | H 1 / C 1
VIN024	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / C 1 | H 1 / C 1
VIN025	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / H 1 | C 1 / C 1
VIN026	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / H 1 | C 1 / C 1
VIN027	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / H 1 | C 1 / C 1
VIN028	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / H 1 | C 1 / C 1
VIN029	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | H 1 / C 1 | C 1 / C 1
VIN030	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | C 1 / C 1
VIN031	2e33	C 1 C 1 C | H 1 / C 2 / - 0 | C 1 / C 1 | C 1 / C 1
VIN032	2e33	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / C 1 | C 1 / C 1
VIN033	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | C 2 / - 0 | H 1 / H 1
VIN034	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | C 2 / - 0 | C 1 / H 1
VIN035	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | C 2 / - 0 | H 1 / C 1
VIN036	32e3	C 1 C 1 C | H 1 / H 1 / H 1 | C 2 / - 0 | C 1 / C 1
VIN037	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | C 2 / - 0 | H 1 / H 1
VIN038	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | C 2 / - 0 | C 1 / H 1
VIN039	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | C 2 / - 0 | H 1 / C 1
VIN040	32e3	C 1 C 1 C | C 1 / H 1 / H 1 | C 2 / - 0 | C 1 / C 1
VIN041	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | H 1 / H 1
VIN042	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | C 1 / H 1
VIN043	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | H 1 / C 1
VIN044	32e3	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | C 1 / C 1
VIN045	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | C 2 / - 0 | H 1 / H 1
VIN046	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | C 2 / - 0 | C 1 / H 1
VIN047	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | C 2 / - 0 | H 1 / C 1
VIN048	32e3	C 1 C 1 C | C 1 / C 1 / C 1 | C 2 / - 0 | C 1 / C 1
VIN049	332e	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 2 / - 0
VIN050	332e	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / H 1 | C 2 / - 0
VIN051	332e	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / C 1 | C 2 / - 0
VIN052	332e	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 2 / - 0
VIN053	332e	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 2 / - 0
VIN054	332e	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 2 / - 0
VIN055	332e	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 2 / - 0
VIN056	332e	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 2 / - 0
VIN057	332e	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 2 / - 0
VIN058	332e	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 2 / - 0
VIN059	332e	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | C 2 / - 0
VIN060	332e	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | C 2 / - 0
VIN061	332e	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 2 / - 0
VIN062	332e	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 2 / - 0

VIN063	332e	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 2 / - 0
VIN064	332e	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 2 / - 0
VIN065	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | H 1 / - 0 | H 1 / H 1
VIN066	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | C 1 / - 0 | H 1 / H 1
VIN067	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | H 1 / - 0 | C 1 / H 1
VIN068	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | C 1 / - 0 | C 1 / H 1
VIN069	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | H 1 / - 0 | H 1 / C 1
VIN070	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | C 1 / - 0 | H 1 / C 1
VIN071	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | H 1 / - 0 | C 1 / C 1
VIN072	2i2i3	C 2 C 1 C | H 1 / H 1 / - 0 | C 1 / - 0 | C 1 / C 1
VIN073	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | H 1 / - 0 | H 1 / H 1
VIN074	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | C 1 / - 0 | H 1 / H 1
VIN075	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | H 1 / - 0 | C 1 / H 1
VIN076	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | C 1 / - 0 | C 1 / H 1
VIN077	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | H 1 / - 0 | H 1 / C 1
VIN078	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | C 1 / - 0 | H 1 / C 1
VIN079	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | H 1 / - 0 | C 1 / C 1
VIN080	2i2i3	C 2 C 1 C | H 1 / C 1 / - 0 | C 1 / - 0 | C 1 / C 1
VIN081	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | H 1 / - 0 | H 1 / H 1
VIN082	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | H 1 / H 1
VIN083	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | H 1 / - 0 | C 1 / H 1
VIN084	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | C 1 / H 1
VIN085	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | H 1 / - 0 | H 1 / C 1
VIN086	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | H 1 / C 1
VIN087	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | H 1 / - 0 | C 1 / C 1
VIN088	2i2i3	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | C 1 / C 1
VIN089	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | H 1 / H 1
VIN090	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | C 1 / - 0 | H 1 / H 1
VIN091	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | C 1 / H 1
VIN092	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | C 1 / - 0 | C 1 / H 1
VIN093	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | H 1 / C 1
VIN094	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | C 1 / - 0 | H 1 / C 1
VIN095	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | C 1 / C 1
VIN096	2i2i3	C 2 C 1 C | C 1 / C 1 / - 0 | C 1 / - 0 | C 1 / C 1
VIN097	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | H 1 / - 0 | H 1 / - 0
VIN098	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | C 1 / - 0 | H 1 / - 0
VIN099	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | H 1 / - 0 | H 1 / - 0
VIN100	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | C 1 / - 0 | H 1 / - 0
VIN101	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | H 1 / - 0 | H 1 / - 0
VIN102	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | C 1 / - 0 | H 1 / - 0
VIN103	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | H 1 / - 0 | H 1 / - 0
VIN104	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | C 1 / - 0 | H 1 / - 0
VIN105	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | H 1 / - 0 | C 1 / - 0
VIN106	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | C 1 / - 0 | C 1 / - 0
VIN107	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | H 1 / - 0 | C 1 / - 0
VIN108	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | C 1 / - 0 | C 1 / - 0
VIN109	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | H 1 / - 0 | C 1 / - 0
VIN110	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | C 1 / - 0 | C 1 / - 0
VIN111	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | H 1 / - 0 | C 1 / - 0
VIN112	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | C 1 / - 0 | C 1 / - 0
VIN113	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | H 1 / - 0 | - 0 / H 1
VIN114	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | C 1 / - 0 | - 0 / H 1
VIN115	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | H 1 / - 0 | - 0 / H 1
VIN116	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | C 1 / - 0 | - 0 / H 1
VIN117	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | H 1 / - 0 | - 0 / H 1
VIN118	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | C 1 / - 0 | - 0 / H 1
VIN119	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | H 1 / - 0 | - 0 / H 1
VIN120	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | C 1 / - 0 | - 0 / H 1
VIN121	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | H 1 / - 0 | - 0 / C 1
VIN122	32i2i	C 1 C 2 C | H 1 / H 1 / H 1 | C 1 / - 0 | - 0 / C 1
VIN123	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | H 1 / - 0 | - 0 / C 1
VIN124	32i2i	C 1 C 2 C | H 1 / C 1 / C 1 | C 1 / - 0 | - 0 / C 1
VIN125	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | H 1 / - 0 | - 0 / C 1
VIN126	32i2i	C 1 C 2 C | C 1 / H 1 / H 1 | C 1 / - 0 | - 0 / C 1
VIN127	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | H 1 / - 0 | - 0 / C 1
VIN128	32i2i	C 1 C 2 C | C 1 / C 1 / C 1 | C 1 / - 0 | - 0 / C 1

4.2. Comparison of Fourier series terms

$$^3J_{\mathrm{CH}} = \sum_{i=0}^{n}\sum_{j=0}^{n} C_{ij}\cos(i\Phi)\cos(j\Psi) + \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\sin(i\Phi)\sin(j\Psi) + \sum_{i=0}^{n}\sum_{j=1}^{n} T_{ij}\cos(i\Phi)\sin(j\Psi) + \sum_{i=1}^{n}\sum_{j=0}^{n} U_{ij}\sin(i\Phi)\cos(j\Psi)$$

Equation A.1

Equation III.1 repeated for clarity. Fourier series used to generate different empirical equations relating $^3J_{CH}$ to the dihedral angle Φ between the coupled ^{1}H and ^{13}C nuclei and the dihedral angle Ψ between an α-substituent and the nucleus the ^{1}H is attached to.[7]

4.2.1. α-substituents

$$^3J_{\mathrm{CH}} = \left(C_{00,0} + C_{01,0}\Delta\chi\right) + \left(C_{10,0} + C_{11,0}\Delta\chi\right)\cos\Phi + \left(C_{20,0} + C_{21,0}\Delta\chi\right)\cos 2\Phi + \Delta\chi\cos\Psi\left[C_{01,1} + C_{11,1}\cos\Phi + C_{21,1}\cos 2\Phi\right]$$

Equation A.2

Equation I.19, repeated for clarity. Equation describing $^3J_{CH}$ in the presence of an α-substituent (X) as a function of two dihedral angles (Φ_{HC}) and (Ψ_{CX}) with a correction for the electronegativity of the substituent relative to hydrogen (Δχ).[7]

Collecting like terms:

$^3J_{\mathrm{CH}} = C_{00,0} + C_{01,0}\Delta\chi$

constants

$+\left(C_{10,0} + C_{11,0}\Delta\chi\right)\cos\Phi$

cosΦ terms

$+\left(C_{20,0} + C_{21,0}\Delta\chi\right)\cos\Phi$

cos2Φ terms

$+\Delta\chi\, C_{01,1}\cos\Psi$

cosΨ terms

$+\Delta\chi C_{11,1}\cos\Psi\cos\Phi$

cosΦ × cosΨ terms

$+\Delta\chi C_{21,1}\cos\Psi\cos 2\Phi$

cos2Φ × cosΨ terms

Therefore, by comparison to the Fourier series, Equation A.1, the constant terms arise from i=0, j=0. The cosΦ terms then arise from i=1, j=0, the cos2Φ terms from i=2, j=0, the cosΨ

terms from i=0, j=1. The mixed terms arise from i=1, j=1 for cosΦ × cosΨ terms and i=2, j=1 for cos2Φ × cosΨ terms.

4.2.1. β and γ-substituents

$$
\begin{aligned}
{}^3J_{\mathrm{CH}} = {} & P_1\cos^2\Phi + P_2\cos\Phi + P_3 \\
& + \sum \Delta\chi_{H\pm}\left(P_4 + P_5\cos^2\left(\xi_i\Phi + P_6\left|\Delta\chi_{H\pm}\right|\right)\right) \\
& + \sum \Delta\chi_{C\pm}\left(P'_4 + P'_5\cos^2\left(\xi_i\Phi + P'_6\left|\Delta\chi_{C\pm}\right|\right)\right)
\end{aligned}
\qquad \text{Equation A.3}
$$

Equation I.20, repeated for clarity. Empirical equation accounting for β and γ-substituent effects for ${}^3J_{\mathrm{CH}}$. $P_{1\text{-}6}$ and $P'_{4\text{-}6}$ are constants. $\Delta\chi$ is the electronegativity of the substituent relative to hydrogen where H indicates γ-substituents and C β-substituents. ξi is 1 or -1 depending on the orientation of the substituent, as defined in Figure I.16.[10]

Equation A.3 can be rearranged as follows for comparison to Equation A.1.

$$
\begin{aligned}
{}^3J_{\mathrm{CH}} = {} & P_3 + P_2\cos\Phi + P_1\cos^2\Phi \\
& + \sum \Delta\chi_{H\pm}\left[P_4 + P_5\cos^2\left(\xi_i\Phi + P_6\left|\Delta\chi_{H\pm}\right|\right)\right] \\
& + \sum \Delta\chi_{C\pm}\left[P'_4 + P'_5\cos^2\left(\xi_i\Phi + P'_6\left|\Delta\chi_{C\pm}\right|\right)\right]
\end{aligned}
$$

Using the trigonometric identity, $\cos(2a) = 2\cos^2 a - 1$

$$
\begin{aligned}
= {} & P_3 - \tfrac{P_1}{2} + P_2\cos\Phi + \tfrac{P_1}{2}\cos 2\Phi \\
& + \sum \Delta\chi_{H\pm}\left[P_4 - \frac{P_5}{2} + \frac{P_5}{2}\cos\left(2\xi_i\Phi + 2P_6\left|\Delta\chi_{H\pm}\right|\right)\right] \\
& + \sum \Delta\chi_{C\pm}\left[P'_4 - \frac{P'_5}{2} + \frac{P'_5}{2}\cos\left(2\xi_i\Phi + 2P'_6\left|\Delta\chi_{C\pm}\right|\right)\right]
\end{aligned}
$$

Using the trigonometric identity, $\cos(a+b) = \cos a\cos b - \sin a\sin b$

$$
\begin{aligned}
= {} & P_3 - \tfrac{P_1}{2} + P_2\cos\Phi + \tfrac{P_1}{2}\cos 2\Phi \\
& + \sum \Delta\chi_{H\pm}\left[P_4 - \frac{P_5}{2}\right] + \sum \Delta\chi_{C\pm}\left[P'_4 - \frac{P'_5}{2}\right] \\
& + \sum \frac{\Delta\chi_{H\pm}P_5}{2}\left[\cos(2\xi_i\Phi)\cos\left(2P_6\left|\Delta\chi_{H\pm}\right|\right) - \sin(2\xi_i\Phi)\sin\left(2P_6\left|\Delta\chi_{H\pm}\right|\right)\right] \\
& + \sum \frac{P'_5}{2}\left[\cos(2\xi_i\Phi)\cos\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right) - \sin(2\xi_i\Phi)\sin\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right)\right]
\end{aligned}
$$

ξ_i takes values of 1 or -1:

$$
= P_3 - \tfrac{P_1}{2} + P_2\cos\Phi + \tfrac{P_1}{2}\cos 2\Phi
$$

$$+\sum \Delta\chi_{H\pm}\left[P_4-\frac{P_5}{2}\right]+\sum \Delta\chi_{C\pm}\left[P'_4-\frac{P'_5}{2}\right]$$

$$+\sum \frac{\Delta\chi_{H\pm}P_5}{2}\left[\cos(2\Phi)\cos\left(2P_6\left|\Delta\chi_{H\pm}\right|\right)-\xi_i\sin(2\Phi)\sin\left(2P_6\left|\Delta\chi_{H\pm}\right|\right)\right]$$

$$+\sum \frac{P'_5}{2}\left[\cos(2\Phi)\cos\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right)-\xi_i\sin(2\Phi)\sin\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right)\right]$$

Collecting like terms:

$$= \quad P_3-\tfrac{P_1}{2}+\sum \Delta\chi_{H\pm}\left[P_4-\tfrac{P_5}{2}\right]+\sum \Delta\chi_{C\pm}\left[P'_4-\tfrac{P'_5}{2}\right]$$

constants

$$+P_2\cos\Phi$$

cosΦ terms

$$+\left(\frac{P_1}{2}+\sum \frac{\Delta\chi_{H\pm}P_5}{2}\cos\left(2P_6\left|\Delta\chi_{H\pm}\right|\right)+\sum \frac{\Delta\chi_{C\pm}P'_5}{2}\cos\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right)\right)\cos 2\varphi$$

cosΦ terms

$$-\left(\sum \xi_i\frac{\Delta\chi_{H\pm}P_5}{2}\sin\left(2P_6\left|\Delta\chi_{H\pm}\right|\right)+\sum \xi_i\frac{\Delta\chi_{C\pm}P'_5}{2}\sin\left(2P'_6\left|\Delta\chi_{C\pm}\right|\right)\right)\sin 2\Phi$$

sin2Φ terms

Therefore, by comparison to the Fourier series, Equation A.1, the constant terms arise from i=0, j=0. The cosΦ terms then arise from i=1, j=0 and both the cos2Φ and sin2Φ from i=2, j=0.

4.3. Coefficients for equations dependent on dihedral angle (Φ, Ψ and/or μ)

The full set of coefficients for Equation III.1, Equation IV.1, Equation IV.3, Equation IV.4, Equation IV.5 and Equation IV.6 are attached in a digital format (CD) and will be made available at data.bris.ac.uk.

4.4. Coefficients for equations dependent on dihedral angle (Φ and Ψ) and bond angle ($\theta_{1\text{-}3}$)

The full set of coefficients for Equation V.4 and Equation V.5 are attached in a digital format (CD) and will be made available at data.bris.ac.uk.

Appendix 5. Experimentally measured and empirically predicted $^3J_{CH}$

Table A.20 Comparison of experimentally measured $^3J_{CH}$ (full details in Chapter V) with DFT calculated $^3J_{CH}$ (mPW1PW91/6-311G(d,p), IEFPCM(chloroform)) and empirically predicted $^3J_{CH}$ for three molecules: 2-ethyl-1-indanone [1], Camphor [2] and Strychnine [3].

Molecule	C	H	Exp. $^3J_{CH}$ (SD) /Hz	DFT $^3J_{CH}$ +6%	Structural Identifier	Fragment Index	Fragment-based approach (uncorrected)	Aydin *et al*. Equation V.2	Palermo *et al*. Equation V.3
[1]	C1	H3	7.9 (0.4)	7.8	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C10	H1	2.4 (0.7)	2.8	C 1 C 1.5 C | C 1 / O 2 / - 0 | C 1.5 / - 0 | C 1.5 / - 0				
[1]	C10	H6a	3.7 (0.1)	4.6	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | H 1 / C 1	CA057	5.6	4.7	3.2
[1]	C10	H6b	1.9	0.4	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / H 1 | C 1 / H 1	CA060	0.1	1.5	0.8
[1]	C10	H8a	2.5	2.2	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | C 1 / H 1	CA059	2.3	2.3	2.2
[1]	C10	H8b	4.0 (0.1)	3.5	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / C 1 | H 1 / C 1	CA056	3.4	4.5	4.0
[1]	C11	H2	7.4 (0.5)	7.4	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C11	H4	6.6 (0.3)	5.9	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | C 1 / - 0 | - 0 / C 1.5				
[1]	C11	H6a	2.9 (0.1)	2.7	C 1.5 C 1 C | C 1.5 / C 1 / - 0 | C 1.5 / - 0 | C 1 / H 1				
[1]	C11	H6b	2.7 (0.1)	2.3	C 1.5 C 1 C | C 1.5 / C 1 / - 0 | C 1.5 / - 0 | H 1 / C 1				
[1]	C11	H7	1.0	0.6	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | O 2 / - 0 | C 1 / C 1			1.2	
[1]	C2	H4	7.3 (0.0)	6.9	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C3	H1	7.7 (0.1)	7.3	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C4	H2	7.8 (0.1)	7.4	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C4	H6a	2.6 (0.5)	2.4	C 1.5 C 1 C | C 1.5 / H 1 / - 0 | C 1.5 / - 0 | C 1 / H 1				
[1]	C4	H6b	2.2 (0.2)	2.1	C 1.5 C 1 C | C 1.5 / H 1 / - 0 | C 1.5 / - 0 | H 1 / C 1				
[1]	C5	H1	6.4 (0.1)	6.0	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | C 1 / - 0 | - 0 / C 1.5				
[1]	C5	H3	7.5 (1.0)	8.0	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[1]	C5	H7	2.2 (0.2)	1.0	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | H 1 / H 1 | C 1 / C 1			1.0	0.4
[1]	C6	H4	3.3 (0.4)	3.1	C 1 C 1.5 C | C 1 / H 1 / H 1 | C 1.5 / - 0 | C 1.5 / - 0				
[1]	C6	H8a	5.1 (0.2)	4.5	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1	HC048	5.0	4.1	3.7
[1]	C6	H8b	5.9 (0.3)	6.3	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1	HC047	6.2	6.8	5.5
[1]	C7	H9	5.0 (0.2)	4.9	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / H 1	HC066	4.7	4.5	4.2
[1]	C8	H6a	2.0 (0.4)	0.7	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1	HC038	1.2	1.0	0.5
[1]	C8	H6b	5.1 (0.0)	5.0	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1	HC039	7.4	7.0	7.4
[1]	C9	H7	5.1 (0.1)	4.8	C 1 C 1 C | H 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1	HC010	3.8	4.1	3.3
[2]	C1	H3a	9.3	9.0	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | H 1 / C 1	HC038	8.8	9.2	7.3
[2]	C1	H3b	5.2	5.6	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / H 1	HC039	2.9	3.2	2.8
[2]	C10	H2	1.4 (0.7)	0.9	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1	HC037	2.0	1.9	1.2
[2]	C10	H9	5.0 (0.1)	4.8	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1	HC007	4.6	4.6	4.0
[2]	C2	H10	3.8	3.6	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1	HC016	4.0	4.5	3.9
[2]	C2	H4a	1.6	1.3	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1	HC023	2.0	3.1	1.7
[2]	C2	H4b	3.2	1.6	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1	HC024	1.8	2.7	1.8

[2]	C2	H9	4.1	3.9	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1	HC016	4.0	4.5	3.9
[2]	C3	H1a	8.6 (0.1)	8.4	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | C 1 / H 1	HC048	8.8	9.2	7.1
[2]	C3	H1b	5.2 (0)	5.6	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / H 1 | H 1 / C 1	HC047	3.3	3.6	3.8
[2]	C4	H2	6.9 (0.1)	6.6	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / H 1 | C 1 / C 1	HC022	9.1	8.6	6.9
[2]	C4	H6	4.4 (0.1)	4.3	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1	HC013	4.8	4.5	3.9
[2]	C5	H10	3.6 (0.1)	3.5	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1	HC019	3.4	4.5	4.0
[2]	C5	H1a		0.3	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | H 1 / C 1	CA015	0.1	2.9	
[2]	C5	H1b		0.8	C 1 C 1 C | C 1 / C 1 / C 1 | O 2 / - 0 | C 1 / H 1	CA014	0.6	3.2	
[2]	C5	H2	4.5 (0.1)	4.3	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / C 1	HC064	6.3	9.4	6.6
[2]	C5	H3a		1.1	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | C 1 / H 1	HC027	1.3	2.5	1.7
[2]	C5	H3b	2.1	1.8	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / C 1	HC026	2.5	3.3	1.9
[2]	C5	H9	3.4 (0.0)	3.3	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1	HC019	3.4	4.5	3.9
[2]	C6	H4a	3.0 (0.1)	3.2	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / H 1	HC036	4.2	4.2	4.4
[2]	C6	H4b		0.2	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / C 1	HC035	0.7	0.8	0.5
[2]	C7	H2	8.5 (0.1)	8.5	C 1 C 1 C | C 1 / O 2 / - 0 | H 1 / H 1 | C 1 / C 1	CA046	10.5	8.4	6.0
[2]	C7	H4a	7.9 (0.2)	7.3	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | C 1 / H 1	CA061	8.1	9.0	6.6
[2]	C7	H4b	4.3 (0.3)	5.1	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | H 1 / C 1	CA058	2.9	3.2	3.4
[2]	C7	H6	4.1 (0.1)	3.9	C 1 C 1 C | C 1 / O 2 / - 0 | C 1 / C 1 | H 1 / H 1	CA043	3.6	4.5	4.0
[2]	C8	H1a	0.8 (0.3)	0.6	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | C 1 / H 1	HC045	0.0	0.5	0.1
[2]	C8	H1b	8.6 (0.0)	8.6	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / C 1 | H 1 / C 1	HC044	5.6	7.8	5.7
[2]	C8	H3a		0.3	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1	HC053	0.0	0.5	0.2
[2]	C8	H3b	6.3 (0.3)	6.3	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1	HC054	6.0	8.2	6.4
[2]	C8	H4a		0.4	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | C 1 / H 1	HC063	0.1	0.5	0.1
[2]	C8	H4b	6.4 (0.2)	6.0	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / C 1	HC062	5.1	7.8	5.5
[2]	C8	H6	3.3 (0.0)	3.1	C 1 C 1 C | C 1 / C 1 / C 1 | C 1 / C 1 | H 1 / H 1	HC019	3.4	4.5	3.9
[2]	C9	H10	4.9 (0.2)	4.8	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | H 1 / H 1	HC007	4.6	4.6	4.0
[2]	C9	H2	1.7 (0.7)	1.3	C 1 C 1 C | H 1 / H 1 / H 1 | C 1 / C 1 | C 1 / C 1	HC037	2.0	1.8	2.1
[3]	C1	H03	8.9 (0.0)	8.5	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C10	H08	0.6	0.1	C 1 N 1 C | C 1 / O 2 / - 0 | C 1 / LP 0 | C 1 / C 1				
[3]	C10	H12	2.0 (0.1)	1.5	C 1 C 1 C | N 1 / O 2 / - 0 | H 1 / H 1 | O 1 / C 1			1.7	0.8
[3]	C11	H13	0.6	0.7	C 1 C 1 C | C 1 / H 1 / H 1 | O 1 / H 1 | C 1 / C 1	HCO148	0.3	1.1	0.1
[3]	C12	H08	5.6 (0.0)	6.0	C 1 C 1 C | H 1 / C 1 / O 1 | C 1 / H 1 | N 1 / C 1			3.8	3.5
[3]	C12	H14		1.8	C 1 C 1 C | H 1 / C 1 / O 1 | H 1 / C 1 | C 1 / C 1	HCO113	2.6	2.5	1.7
[3]	C12	H23a	6.1	6.2	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | H 1 / C 1	HOC014	4.9		
[3]	C12	H23b	8.3	8.5	C 1 O 1 C | H 1 / C 1 / C 1 | LP 0 / LP 0 | C 1 / H 1	HOC013	7.6		
[3]	C13	H11a	3.9	3.8	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | C 1 / H 1	HCO134	4.8	4.9	4.2
[3]	C13	H11b	2.0	0.1	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / O 1 | H 1 / C 1	HCO125	0.2	1.7	0.7
[3]	C13	H15a	3.5 (0.1)	3.8	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | H 1 / C 1	HC050	2.9	3.2	3.3
[3]	C13	H15b	7.9 (0.0)	7.4	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / H 1	HC051	8.8	9.1	7.1
[3]	C14	H08		0.3	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / C 1 | N 1 / C 1			0.4	0.2
[3]	C14	H12	1.8	1.1	C 1 C 1 C | H 1 / C 1 / C 1 | C 1 / H 1 | C 1 / O 1	HCO176	1.8	1.4	1.1
[3]	C14	H16	6.4 (0.1)	6.1	C 1 C 1 C | H 1 / C 1 / C 1 | H 1 / H 1 | N 1 / C 1			9.4	6.3
[3]	C14	H20a		1.4	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | H 1 / N 1			1.2	

[3]	C14	H20b	5.4 (0.0)	5.0	C 1 C 1 C | H 1 / C 1 / C 1 | C 2 / - 0 | N 1 / H 1			9.3	
[3]	C14	H22	8.7	8.5	C 1 C 2 C | H 1 / C 1 / C 1 | C 1 / - 0 | - 0 / C 1	VIN124	8.0		
[3]	C15	H13	3.3 (0.1)	3.8	C 1 C 1 C | C 1 / H 1 / H 1 | H 1 / C 1 | C 1 / C 1	HC040	2.3	2.4	1.6
[3]	C16	H08		0.8	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | C 1 / N 1			0.8	0.3
[3]	C16	H14		6.2	C 1 C 1 C | H 1 / N 1 / C 1 | H 1 / H 1 | C 1 / C 1			9.3	7.0
[3]	C16	H17a		0.9	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | C 1 / H 1			0.7	0.4
[3]	C16	H17b		6.1	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | H 1 / C 1			8.7	6.3
[3]	C16	H18a	6.3	5.9	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / C 1				
[3]	C16	H18b		0.3	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / H 1				
[3]	C16	H20a		0.1	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | C 1 / H 1				
[3]	C16	H20b	6.9	6.8	C 1 N 1 C | H 1 / C 1 / C 1 | C 1 / LP 0 | H 1 / C 1				
[3]	C17	H08	5.9 (0.1)	6.3	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / N 1			7.3	5.8
[3]	C17	H16		0.7	C 1 C 1 C | C 1 / H 1 / H 1 | C 1 / C 1 | C 1 / N 1			0.5	0.3
[3]	C18	H16		1.0	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / C 1				
[3]	C18	H20a	9.4 (0.1)	8.9	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / H 1				
[3]	C18	H20b	3.4 (0.1)	3.6	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / C 1				
[3]	C2	H04	7.6 (0.0)	7.4	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C20	H14		1.6	C 1 C 1 C | N 1 / H 1 / H 1 | C 2 / - 0 | C 1 / C 1			2.9	
[3]	C20	H16	3.6	3.0	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / C 1				
[3]	C20	H18a		0.4	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | H 1 / C 1				
[3]	C20	H18b	7.5 (0)	7.7	C 1 N 1 C | C 1 / H 1 / H 1 | C 1 / LP 0 | C 1 / H 1				
[3]	C20	H22	5.9	6.3	C 1 C 2 C | N 1 / H 1 / H 1 | C 1 / - 0 | C 1 / - 0				
[3]	C21	H13	7.6 (0)	7.2	C 1 C 1 C | C 1 / C 2 / - 0 | C 1 / H 1 | C 1 / C 1	VIN028	7.0	9.4	6.9
[3]	C21	H15a	6.5 (0.1)	6.1	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | H 1 / C 1	VIN022	9.0	9.3	7.2
[3]	C21	H15b	1.9	1.1	C 1 C 1 C | C 1 / C 2 / - 0 | H 1 / C 1 | C 1 / H 1	VIN014	1.2	1.4	1.0
[3]	C21	H23a	4.6	3.9	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | O 1 / H 1				
[3]	C21	H23b	8.4	8.1	C 2 C 1 C | C 1 / C 1 / - 0 | H 1 / - 0 | H 1 / O 1				
[3]	C22	H14	3.6 (0.3)	3.8	C 2 C 1 C | H 1 / C 1 / - 0 | C 1 / - 0 | C 1 / C 1	VIN080	3.9		
[3]	C22	H20a	5.4 (0.0)	4.8	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | H 1 / N 1				
[3]	C22	H20b	4.7 (0.0)	4.8	C 2 C 1 C | C 1 / H 1 / - 0 | C 1 / - 0 | N 1 / H 1				
[3]	C23	H12	2.7 (0.1)	3.1	C 1 O 1 C | C 1 / H 1 / H 1 | LP 0 / LP 0 | C 1 / C 1	HOC012	2.6		
[3]	C3	H01		7.2	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C4	H02	7.8 (0.2)	7.6	C 1.5 C 1.5 C | H 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C5	H01	8.0 (0.0)	7.9	C 1.5 C 1.5 C | N 1 / C 1.5 / - 0 | C 1 / - 0 | - 0 / C 1.5				
[3]	C5	H03	9.9 (0.2)	9.7	C 1.5 C 1.5 C | N 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C5	H08	3.2 (0.1)	3.0	C 1 N 1 C | C 1.5 / C 1.5 / - 0 | C 1 / LP 0 | C 1 / C 1				
[3]	C6	H02	7.9	7.7	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | H 1 / - 0 | - 0 / C 1.5				
[3]	C6	H04	5.7 (0.0)	5.5	C 1.5 C 1.5 C | C 1 / C 1.5 / - 0 | N 1 / - 0 | - 0 / C 1.5				
[3]	C6	H08	3.8 (0.0)	3.5	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | C 1 / C 1 | C 1 / N 1			4.8	2.7
[3]	C6	H16	2.5 (0.2)	2.6	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | C 1 / C 1 | C 1 / N 1			4.9	4.0
[3]	C6	H17a		3.1	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | C 1 / C 1 | C 1 / H 1			3.4	3.7
[3]	C6	H17b		0.5	C 1 C 1 C | C 1.5 / C 1.5 / - 0 | C 1 / C 1 | H 1 / C 1			1.0	0.7
[3]	C7	H01	2.4	2.5	C 1 C 1.5 C | C 1 / C 1 / C 1 | C 1.5 / - 0 | C 1.5 / - 0				

[3]	C7	H13		0.6	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / N 1 | C 1 / C 1	0.7	0.2
[3]	C7	H15a	1.7	1.4	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / N 1 | C 1 / H 1	1.9	1.0
[3]	C7	H15b	7.3 (0.0)	6.9	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / N 1 | H 1 / C 1	9.4	8.0
[3]	C7	H18a	5.1 (0.0)	4.7	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | N 1 / H 1	8.0	5.5
[3]	C7	H18b		0.1	C 1 C 1 C | C 1 / C 1 / C 1 | H 1 / H 1 | H 1 / N 1	0.5	0.2
[3]	C8	H12	5.8	5.6	C 1 C 1 C | H 1 / N 1 / C 1 | H 1 / C 1 | C 1 / O 1	8.7	5.0
[3]	C8	H14		6.5	C 1 C 1 C | H 1 / N 1 / C 1 | C 1 / H 1 | C 1 / C 1	9.4	6.8
[3]	C8	H16	6.9	6.5	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | C 1 / N 1	8.1	4.8
[3]	C8	H17a		8.0	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | C 1 / H 1	8.6	6.3
[3]	C8	H17b		4.3	C 1 C 1 C | H 1 / C 1 / N 1 | C 1 / C 1 | H 1 / C 1	4.2	4.4

Appendix 6. CSD datasets

6.1. CSD reference codes for Dataset 1 (387 structures)

ACAVIJ AFIMUZ AHOWEZ AMNAND AVESAV AVESEZ AXELUP AXICAM AXUQEQ AYAJOA AZOKAB AZOKEF AZOKIJ BABJUJ BACCHO10 BACCHO11 BACCHO12 BADZAJ BAKHOK BAPOCM10 BEGKIJ BEJXOD BEMTAQ BEVZOR BEWNOG BEYDOA BIDKUV BIFQIR BIFQUD BOBKEI BOCHIK BOGKIR BOGZON BOPVIN BOTJAW BOTJEA BUCCUX BUKKEX10 BUSFEA BUSFEA01 BUVRAN CAVHIR CAWPIZ CAYHEQ CEDWOZ CERCEH CEYVAD CICRIQ CIZZOB CMPANL01 COBXOH COBXUN COBYAU COCCAZ COYSOY CUBYAZ01 CULDAO CUQMOS CUQNAF CUVPUE DAJLIJ DAZPIE DENPUH DEZDAN DIHVEV DIJWUP DIJXAW DOHDUZ DORBUJ DOVGAX DOVGEB DUDKUJ DUKKAV DULWAI DUTXUM DXTMPO EADSUX EBIBEW ECADIW01 EDEGOK EDIFAZ EFIGEG EGUFOD EGUFUJ EMESIZ EPFRED02 FAFJUR FASDAE FATRUN FAWLEU FEVCAL FIGLOW FIQTUV FITVOT FOBQOC FOGHAL FOHRAW FOLVUX FONYUD FOQKUT FOXCUR FUFGAO FUGTOQ FUGTOQ01 FUTZOJ FUVVAT FUYNUI GACLII GACLOO GACLUU GAGQAG GARGAJ GATVAY GAYCOY GAYKAS GAYKEW GEFMOT GEJLAK GEKRAR GETBEM GETCUE GEVNAX GIMJAP GIXNIM GOGPUP GOGQAW GOGQEA GOKNEZ GOKXIN GUXGUB GUYFIR HADJEC HAVVUW HAVZEJ HERRUR HERSAY HEXSAF HIFCII HOTJAB HUCTAB HULXIX HUXQAU IFAHAZ ISELEX JAGQOX JASSOL JASWUV JAYVOU JAZTIN JESGIX JESGOD JIBBOL JIRKOL JOLBIW JOQPIQ JUFQAE JUVFAH01 JUWDOU KAGNAJ KAGZUN KANDUY KANJOZ KANJUF KEPSON KEWREJ KICDEH KICDIL KINSEF KINSUV KONWAL KOPKUV KOWBUV KOWCAC KOWDAD KUKJOR KUMTAP KUMTIX KUTMIW KUZKAS LADNIP LAKGIO LAQYUY LEBWEV LEFHAG LEPJAS10 LEPYUA LIBMIU LICKAK LIGCAH LITQIP LIXMIP LOLPAF LONPEJ LONPIN LUDTUA LUDWAJ MANGOY MAXJUQ MAZCIB MBLVBT MIDYAA MIRMOQ MOFYOX MOHZOA MOZLIX MPASSF10 MUFXUG NABHUU NAFWEW NAHYOK NAMVUS NEMXUY NEXHUV NEXNIN NEXNOT NEZFON NIWZAV NOBYEK NOKVUE NOKWEP01 NUCXOA NURXED OCAXUM OCILER OCUWAK OFOQIL OGITIH OGITUT OGIVAB OJIXIP OKAJIU OKITIL OLESUU OMULIS OMULUE OPELEB OSUVEE OSUVOO PABBEA PABBIE PABDAY PABFII PAPGAO PEFCAE PENCAO PIMPAE PIZFIP PIZFOV PIZFUB POPZAX QAJSIF QAQHIB QEMQUU QIQCIC QUMLEQ QUQYAD QUYTEL QUYTIP RALGER RALGIV RAMGOC REXMEO RHSLCT10 RIMGUR RIWCIK ROJSIU ROJSOA ROJSUG RULGOU RULGUA RULHIP RULNOB RUSYOT RUSYUZ SAFBEH SAFBIL SAFBOR SASLEF SEKZUD SEMKOK SIFFET SIQGUU SLSPRM SOQKOZ SORROH SORRUN SPEREK10 STMDIN SUZBOF TAGTOK TAYTIY TEGZOU TEMKAZ TESLOT THYSNL TIFNUS TOQBUY TOTNAR TOTNEV TUHWAW

TUQZEM TUQZOW UFEHIW UFUCIJ UFUQAO UJISAH UPAXOZ URACOG URENOV VABNER VATGEE VEFNOJ VEKMED VEPBEX VIWGIS VOHFUU VOKZAX VOLVOH VUFREU VUGWUP VUKCUZ WABKAL WADGIT WAKYUE WINPUG WIXWEG WIYKEW WOBKAZ WOBKIH WOMSUM WORYIL WORYOR WOVFIY WOXKEZ WUDKEN WUDKIR WUDKOX WUFMEP WULYEI WUTVUE XAHTIM XAZWUQ XAZXAX XAZXEB XEBMIB XEBMIB01 XEBMIB02 XEBMUN XEBNIC XEDKUN XEJNUX XENQOX XEPGUW XEPHAD XEVJEN XEVJEN01 XICPAB XIZDAM XONXON XUKNEY YACNEV YAJZEP YIRLOC YIRQIA YOLLES YOLLIW YOLLOC YONQUP YOTLAV YOTPON YUQRAE ZAYZIK ZAYZOQ ZEPTOF ZEPTUL ZEQCOP ZERCUW ZERJOX ZETNAP ZIHHAZ ZINRUJ ZOPPUR ZOPQAY ZOXHUP ZUFPUL ZUMBEO ZZZQAI02

6.2. CSD reference codes for Dataset 2 (1,002 structures)

ZIJQEQ FEYCAN NXENAM FAGQUA NIBZIH KALPUK SUZVIT IMTHZN IRAQEZ FIVXIT DUZLUG COBQUF QIBLUI ZUFQAU DIDHIH GUNWUJ HIRMAX CADSEF VOLLEO EQEHEP SUZRAG ROGNAD PODGAR NABLEI HUFDIW DUDCUC02 HOHVIL LEGJOY HUTKIR AGIKUY VUQTIM NPIMPY ECADEU HEJFEJ NAYMIJ LUYPAW QOVTEB XIGGAX QOHSIR WECWUW JEVYOZ QIJYOX ANUDEU BOSKEA DUYYED BEGREM JIPFAQ YERDAA CIBCUL10 MUBQUW YAFNOL LIQBUI PETYAO BACVUX WIGRUB KUSPIY QUJYOL GEBKEF WIFNIJ FUNCAU ICAPOR01 MBANOX10 ZEWQUN COCFAC XAFFIW GIRJIC CUSJEG GEZLIG COKYEI KIJPEZ IZIREQ JODYEG MOHTIN IYIKUY NAXXAM PATWEN01 WAHFOB JEXJAX GUKTUB DOVWAN TADNUI BEQYOL PARGIY LIVWEU CACYIR ASEXOM UMILEH RODVIP OZUYIT WUJVII OXUXIQ CYHXUR08 IFOSEC PIHBEP SOJXIZ JATMAS MAZDAU LOMMEF PADXEX FURZAU QIWDIL OBOCOA PUBXIT VACZII JERVUY MELFIT06 AKILOV QEVSEQ WAFJUL TMPPIO13 REMHIC RUCYAR GIRRAB ASATIZ FONCAN ABEJOH ITIJOL01 OCUSUA04 SAQKAX LENTEE ETIMQO10 BAQYAT TEWXAW SOHFEB HALSAO YIHKIK MOXHIS VEJQEH OGUGOM EVIBIV NIQCUM02 MUXLEW HAVJUJ JEDAEY WOCJAZ INUZOG YOGWOI GUHFEU BIDDUN JOPCIA KOWBUV XAQSUE HESZIP OPIBEV VAGVIK ROGGUQ LEMCIQ IRINAY WIKBEX NEQLAX ZAPWIW MAYVUE UDAZOQ NAHSIZ TITJAI EBOKUE RIBZUY BUDKOB MOCPEB MEHLOA BUVDON LIPSIN WUGLEP DECQEH ECUJUH RIRLUA GELZUS SUVQOP FADGUO TAPPIK AXBZOP UKEMON HUKDOH RULHUD PIKZEO PEYMIQ09 XOZJIF RUPVIJ VEVMOA MANDEM MAHSIZ LAGLEK OLAHUE HUHLEC KUKZEW DAHYET BAYSOJ ZEGGEZ QEGWEG01 EXIMEE MEBTIY VOVHET WILWET DAMPCX LAHKAH RUGHIM ZERJOV NUJFII KELKOB MEHXEE ACOQIS EFEQUD AEBDOD10 OQAXAG WATTEQ OREDAS ZULHIX10 LAXWOW JAXKID OHCHRY01 COWBIB IJAMUD POHYAN ZAJWAI01 GANHOS DABWUY10 TUNZIN EMODIU HEGROA IPAWIG MAJWOK

NAQWEH DOJWIJ TOCWAJ DEXWAF CUJLAW SINMIM ODEMEQ SAWMOU SAWDUQ FEKMAL VUSLEB EGOTUQ GUXKUH ISISUY GEMGEM POSJEN WOKNOZ NEGCIL MEVZES SUTHAZ07 BILQEU01 YUXJAE JUPMUC QUTLIA RUVXEN DATSEY AYUWIB PLBULC MAZNIK AJAKUT FUNZEU UCUYAT PIYTIA FUFTAC EMUYER UWIDOU SUKKIS BULQIK QAJBIN JATBOW02 LEKKAN ZZZEIY01 DAFMEE QETYOE GIBREO VUFSEV NERTOU BORLIF DADBIV OLIBIV FIGFAE ARPLYS HAFKEH NAHMIS TAMNUS SECTAX MUJBIE LITPIO NUNHEK MELXAC KOHHIZ BIVTAR CIDPOW IDUZUD TIYJEQ01 UVUXOA GUYNIX RESROW FEMPAP01 TUWKED YIVXIK LOBHOZ ATOYUD CODQAP ZUSYIV OBUSEJ RUGPOZ ZUYKUB VEWMAM JAFCOI NATMEA EKONIB PURSEB01 RIZFEN UKOMAK VOFYEV TAZXOI EVATOK KEQPEB SOCJID BIJKOV AWETIH TERDUQ PODHEW IRERUS VENROX WEHKOK AYAROI LYXPYR MOXXII HACHEA AMEWAS ROBYIQ HAKHEJ YAXTUO YAGMIE MUMMIS LIXQOZ UHEYOV ETOZES MITJUW MAMNIZ FUYLIV JEHBEE HUKLEE ELAKEH RENTOT KEFRIW QECTAV YERRUI02 KUHTEO WOCMUY01 JUYGAL SAGTOM TOLJEL JOLPEH UNIDIF GIHTIA XANTOX KARYAD XOSWAF IGATIU BOCJAE HOTHED SEVYIC JIQZIT NOZXUV FIZNOS01 AZIHEW ONOTAO HAGLUZ OTOGOV FUTMUC BIBGUQ EVICUI RAYTEQ WALFEV IQIQOQ ZIWZIQ CANPUC HOXBSA01 IQANUL RULJAK KUKBUO HAWDIT XIVYEG FIHYIG QAWQEM VAWMUD KAKRIA WOKJOV YOSRAZ RUJCII01 TAMTEG FARQIZ TUKSIC DOMSUT DOMJUM BABBOX MZPROD CAJNAC BONKOF WADSIG CISQIF ZURCEW AZALAP YUHYAB MOTFAD TOCVAK QEZBOM TABDIJ XOGXEX SETGEE OGUCAV TOFHIF PILCOC MAZHAY GOJXOS JOJCOA SAWVIX GUNQUB TEHMUO MUBKOK OJOPIN OFEYUT VAFVUU KOJYOX MOLHIF CUZYEB IYIWIW SULMIW ACIQEI CEZMEB TIMYUL JUWNAQ UBAWEB RIJJOJ TEUVID YABBIP WUSPOQ SEMVUB GOKSUV OYADAV XIZTEF QOHMUX SABVAU SIPJIK MOXYUV DIFYEX PUPFOW CEQYON KUJDOI PIWWOI TAVJOR XIKZOI KIFTOI RASWEN GILFOY TIHTAG MOYKOA JOZGAG VEVZUR10 DIBENZ14 YABCUA FUSJOU LOMTIS TUDYIB WOLWUQ IGAXOE PUWROQ KUWXEG GITKEA MUBSIM QULPEU BALQUB OHADEI EKOJUL INOKEC LUDGIA DEGBOG WAJREH IHARAK FUBWEG TENNOR DORFIB ANUTUZ NEHMOE ELOXOS WAFKUM CANTOD DCLBQN02 NABSER RAPJUO UGOTUH CEZMUQ HUFKOJ EDUJAO KEPPAY WUGSEY ONOYOG WEGPUV YOGYIC GUYBOR01 UDAHUE ILIREA DIFFOO FUZXAB YAWKAK OHULOU OQOFIL DUVHEH10 KEZKEG OFISEC HDMBZD IDEVET ODIRAU PIXNAN SOFHUQ KIRPIM DILFUZ HEVZUD COZYIB QIKHEZ FUWXEB ZEJJIH TOZQIK YIPPAQ METYUG PEXHOR WEXQAS01 AQIWAY WIPJEK KEQXIO MUBBAM EKUWAK XIDMIH01 FACXOX KEWKII NAMMOE IVOHIL SAYNIR YIZNOK ABESOS PUGHUV GUDQED CLDOUR OTIJOS UMEGAW ITELOK FILZOR BULWUC YONTAX FAYWOT GUGZIS ULOGEI ETZAND WUQKIE ZEKQEN VUXBEX

WEFMEA PUJRIW POTRUN TEYDAC01 YAGHUL YERTEV MONTIS03 COTDIA PINCIA FAWJUK WUXHON BAMJEH WOCBEV AQIBIM UBAZOO ASISIG LACBOG VEYPAR HEWTOT OKUCOO TEWLEM MIZDAB ANOXEH SILYIV RUNQIC HUHYUG IJEQUJ PELCOA FIYTEN HEZWOZ SONQAN VIXNIA UHELOI ARABOL TIWFIQ FOCWUR ZILBAX IBUPEC JUBBEP YIGYOE BEFJON UCOQUB XETPET VOBTIR EHAQEJ PODJUO PUJNEO SOGQUA HOYPER LAFFON QAWJEE DUTRIU02 BANFIH NETGEY BUHKEV PACSER NOMQEN RIPQIS IPCUND WOCNIN SOJHUU VIJDID SUYTUC LIVTAL WEJQIM GUQHOR DAMPAL CINWED VEQDOL VUSBIU KAZDIA UCEWIL GIHTAU LAPWIJ HECLAE SERPAI ZENNIP LUVSIE YUYHAD GECQOU TOHVUI LIDBIJ UZENAQ01 XUHPAR QUFYAR SOJCID POTGIO XIFLAA VODNOR MAEPYZ AHUDOY DIFDUS APIPAS AMAYUK PIWCON ZOGRIW ITAZUA OCETIC NUYGOD WUCZAX QIFJUM CADGOF KIZLEL WONQOF MOLMUV OCIFOV HIMMEW DAQTIY ILAJEM LORTUI01 IJUFAV JIZWUK AXEPIC ROCPIJ EWOTIT UDIJIC EDOYOM HACJOK WUNGES MEWRIR IRAVII IFUZAM QIMBIX WAPGOK DIBREL JAHYEW TIKWOB XOPPIC KUFGUP GEFBIE RAZWUK PIQTEQ VOMSIZ LETGEW AXIMEZ IDEGOQ GOKDEQ MIDKOA VAFFEP MEWDAT BANLEI NUNMAK GIDXIA CIRWOR PARXOX EZUKOA ELAJEI DUZREW USOLUK XULMUO WIXWIJ VELSEM MEXPEM IFAFAX PEXKIO FARYOM TEGGET COVPIM SUQVAC FADBOC QUYVAI UHOSEQ TULYEE EVECOZ MAKPET NITSOZ MAJVIF DIGWEX PIHZUB WECXUZ PIBRUP WIGMUW EVUSUK JACDID NAYLUU UHIDIA SATCIZ POJWUG DAVPEW LELWAA PORSIZ OPAHAQ YUJCUC SOYPAY GIMQOJ LAQVEG EMUQEI CIFREP DAQNOY AQAJUZ BELQOZ HUSSEU AXABEH GUSRIV01 DOWHAZ QECXAZ NUPYEE WIYNIB BUNVAH01 GAHTEO XUWMUZ COYWUI IMSTXP10 PEMFIW PIHTOR VOCVIU WULVOQ DEMTEU WUJVON ZZZJFA01 YOFMUC SUHBUT SODZIW XEXQAS TIMGOL FUXQIA JUZTUT PAJWUV PICFAI XIPQUK NOSYOJ01 NAHYOK DATCAD HETJEV HODREY TISRUJ KEZPUC SUTHAZ37 BAHZEP01 JIVMAD EZEGEX GURLOV DLHTDA10 QAJWUV JEYKOO WOBNEI PAMTED WICPOP IGUPEH COLDEM APOWIM ITINEE VUZREO MODMAT POBVIM MOLZES BUFTIG GILHUG CINXAY GOZPEQ UCUVET SUSDER FIFXIB01 HESZAH DAHQOV PUFNIP ZZZSSO03 HEKZII LOHVEJ GENMOC GOHCEN REDDOU QORQIY YAHSAD ZAKMIH02 MIHZOT BUFGOY LUMSET YADVOP DUZHAI SARPUY TERTAM GICVOE QURVOO DINGAI01 PAWTOW01 HAYWOU JOTDIF RIKYUF QOWQEY DAHYAP WOYLUR NIHBEL WUKRAW ROGGOK PEMBAL SEMKOK KOVHAE01 LUJZEW KOGSEE GUGJIC ZASYEZ ZELNAF FESGUG BUHQEA SOVPUO VEKXUE JILKOE01 YICGIA NEFPOF RUJPUJ VIPPAM ESUHOR BUSNUA POYDUD WIFKOL01 GAPTOI FIBBEY DAJSIQ01 OGERAV FOLKUM ERE REY OQEQOS ASEHUB01 POSBIK ELOPEC BUSPEK DOKHAN BAZNEW CUCWIH SEDJEQ SUDJOS POFQAE HIBBUQ JUJTIR HEMYON

CIDFAX SEMDET LIGTIE LIWGAA JULJEF ENSULF03 JOFGER ZUGQUN QAJPOI OQERAE01 QOBSOQ JANYON IPAPUM ZUPFUL VABBUW GIQLIC ZILGAE WOHRES CEMCEE ULATAE YOLYUU KIKSAZ POLMAG MERPUV GIYZUK FABYAL LAXROT NAMVII RELRAD CUHDAM SEZMIT HUDJUN NENFIX01 CAQJEK QUDWAP YOBQAH TUHWUO UWASUI IMONON NUHLEH01 IDIBUT ISIHUN02 FEMXAY QIKYIU BAKVEQ BIMCEG XIXBEM MORYAU CAZMOH TUWJEC JADJEE IFAXAP HEHQIU KAVZIS XUQMIF TUWYOA MELNAT BEQZOO VEFFAP XIQJUD ZAPYIA YEQCEC LEJMOC NIMMUS WUPQUV AZBAPY IMIRON TIDNOK GEGZIC BETPEX DASWOJ JEWCEU BAPBUQ ZESVOK YAHTOQ ZCPTAZ LUMYOH XAZMAN YOJWOL QUXJEA CEGFOL VAHBAK YIGGIF FATVUS XOQLUK PEMFET OPILUV BAMKEF BIVFER MAHRUJ XUYRUF OYONUN VIWGOY AZITEI KEGWOJ SIQHEF IQENIC WEGYAJ FIZJED TANZEN YOPXUX XICGEX APEMOZ ISONIO AWUGAB NAJQIA ISUSAS QUGKAG PULCHB11 ZIJQOY NIDWEE PUSKUL POZGOB RAFYED JIZJIL YABVUT SAWRAL RICKEV POSNAN XUFDAE BEHXES TUQSUV DADXIP OKIKAW AYASEZ ACIVUF

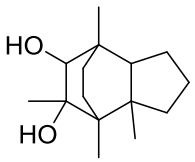
6.3. Bond angle corrections

Table A. 21 Full set of coefficients for Equation VI.1 from each CSD dataset.

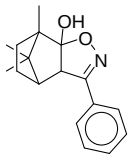
	A_1	A_2	A_3	B_1	B_2	B_3	C
Dataset 1							
HC064 only							0.964
HC064 only				1.04×10^{-2}	2.04×10^{-2}	3.56×10^{-3}	0.967
HC064 only	1.81×10^{-3}	-4.68×10^{-4}	-1.18×10^{-4}	-1.91×10^{-2}	3.39×10^{-2}	2.10×10^{-3}	1.021
All fragments							0.952
All fragments				5.59×10^{-3}	2.46×10^{-2}	1.74×10^{-3}	0.962
All fragments	2.71×10^{-4}	-7.96×10^{-4}	-1.60×10^{-5}	2.26×10^{-3}	1.98×10^{-2}	1.71×10^{-3}	0.965
HC fragments							0.955
HC fragments				6.19×10^{-3}	2.57×10^{-2}	2.13×10^{-3}	0.963
HC fragments	2.90×10^{-4}	-7.15×10^{-4}	-2.32×10^{-5}	2.70×10^{-3}	2.15×10^{-2}	2.29×10^{-3}	0.966
Dataset 2							
All fragments							0.945
All fragments				6.37×10^{-3}	1.86×10^{-2}	1.26×10^{-3}	0.942
All fragments	8.59×10^{-4}	-9.49×10^{-5}	-8.82×10^{-5}	-4.06×10^{-3}	3.38×10^{-2}	4.00×10^{-3}	0.955
HC fragments							0.956
HC fragments				9.03×10^{-3}	1.45×10^{-2}	2.31×10^{-3}	0.935
HC fragments	9.71×10^{-4}	5.53×10^{-4}	-1.76×10^{-4}	-3.59×10^{-3}	4.34×10^{-2}	8.70×10^{-3}	0.955

Appendix 7. CSD datasets

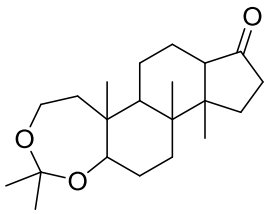
7.1. CSD structures with reference codes for Dataset 1 (387 structures)



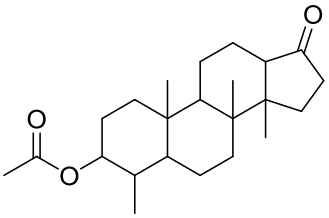
ACAVIJ



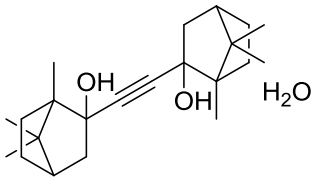
AFIMUZ



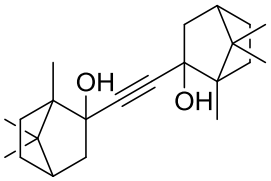
AHOWEZ

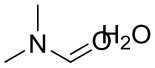


AMNAND

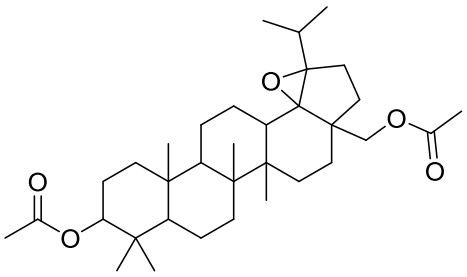


AVESAV

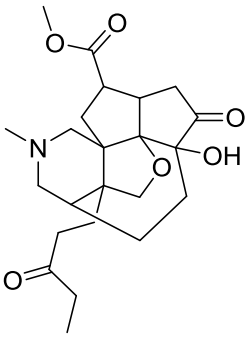




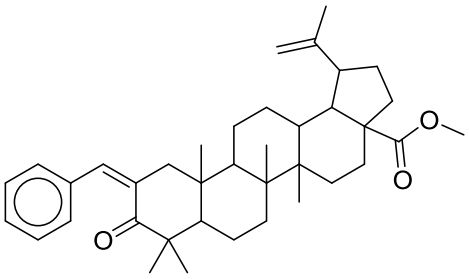
AVESEZ



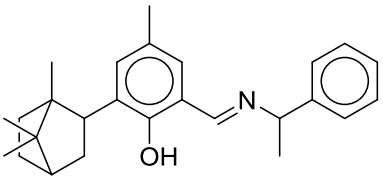
AXELUP



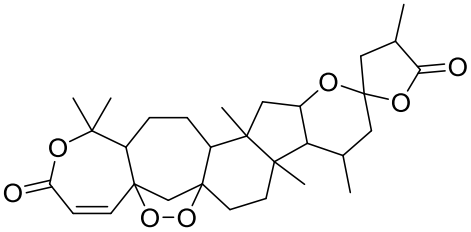
AXICAM



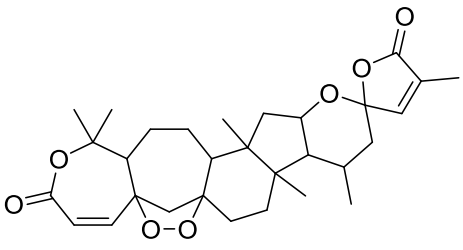
AXUQEQ



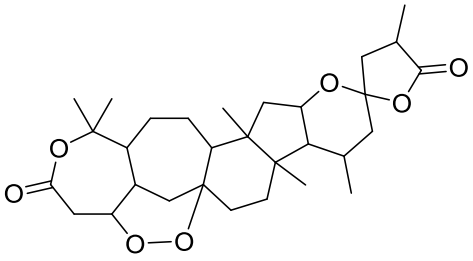
AYAJOA



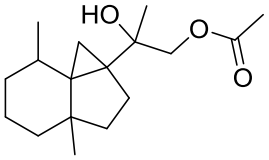
AZOKAB



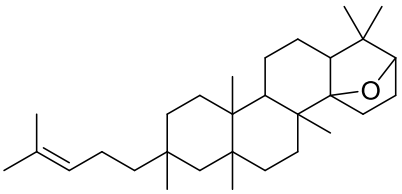
AZOKEF



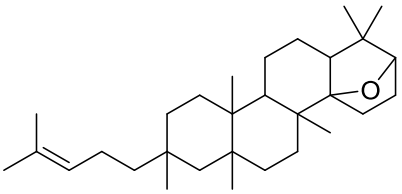
AZOKIJ



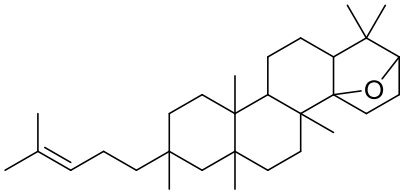
BABJUJ



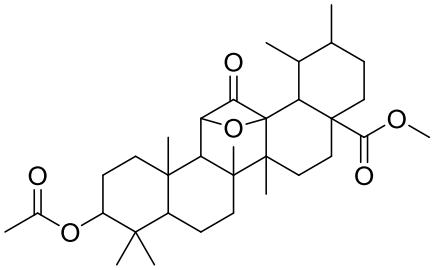
BACCHO10



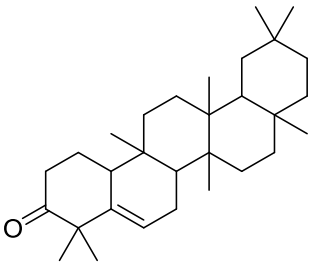
BACCHO11



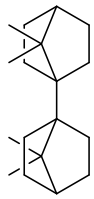
BACCHO12



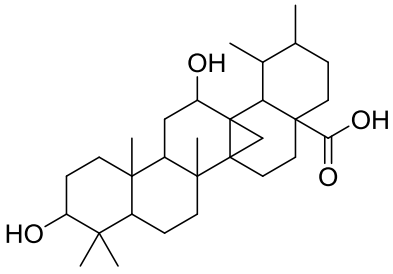
BADZAJ



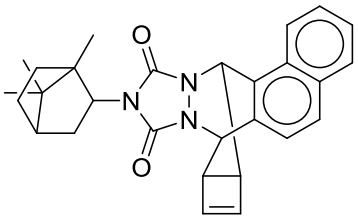
BAKHOK



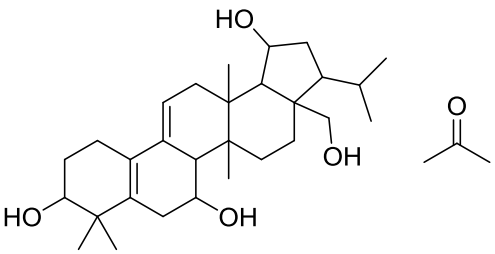
BAPOCM10



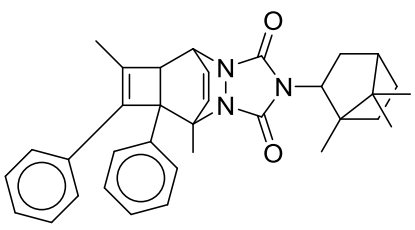
BEGKIJ



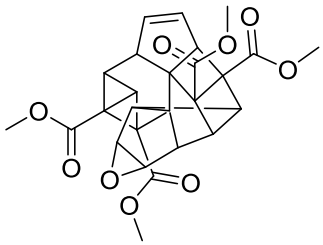
BEJXOD



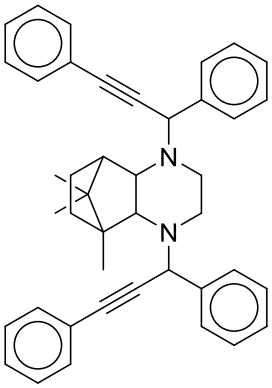
BEMTAQ



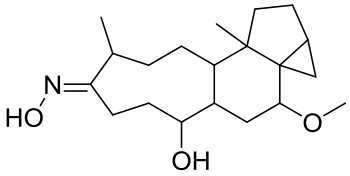
BEVZOR



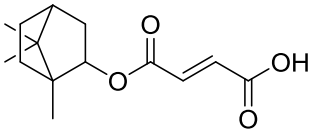
BEWNOG



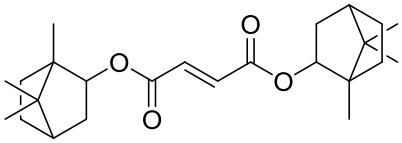
BEYDOA



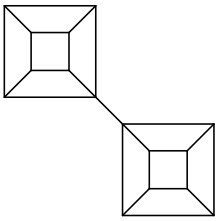
BIDKUV



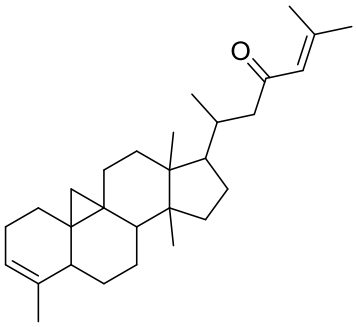
BIFQIR



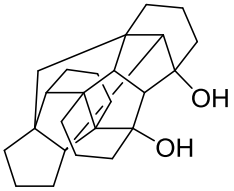
BIFQUD



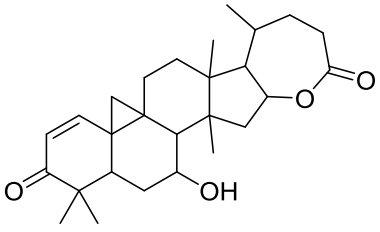
BOBKEI



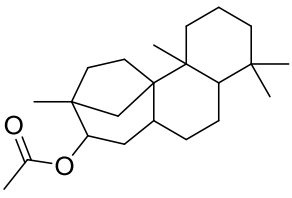
BOCHIK



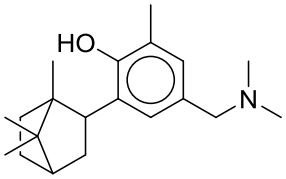
BOGKIR



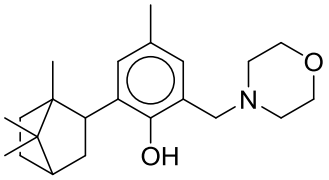
BOGZON



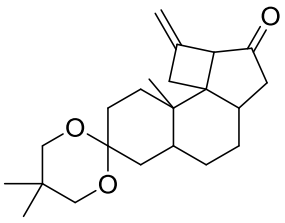
BOPVIN



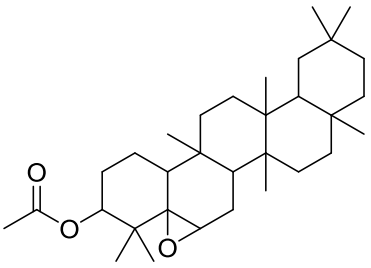
BOTJAW



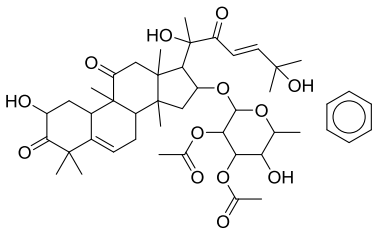
BOTJEA



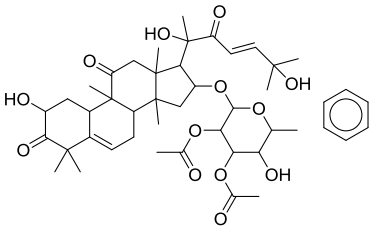
BUCCUX



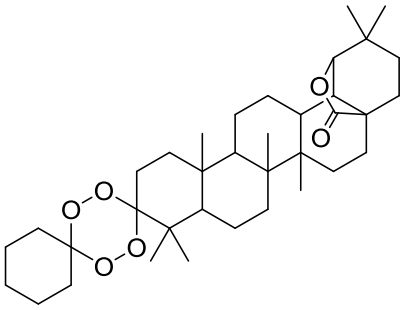
BUKKEX10



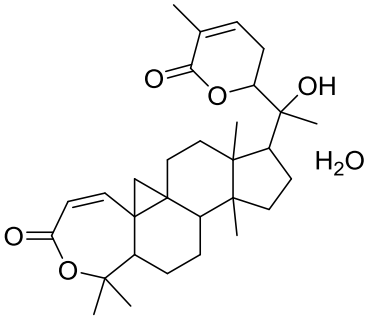
BUSFEA



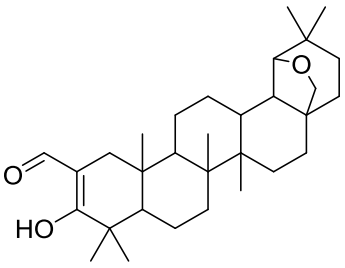
BUSFEA01



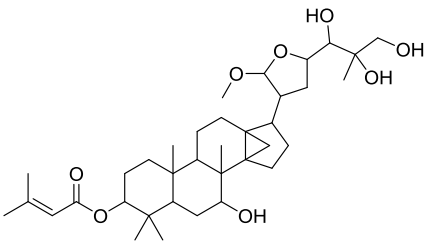
BUVRAN



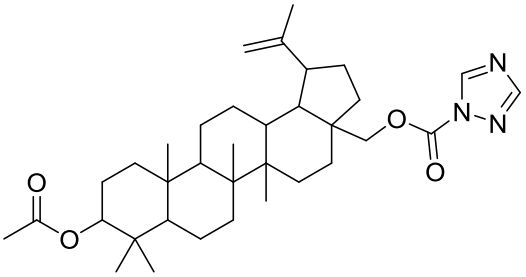
CAVHIR



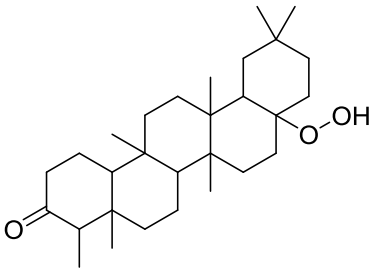
CAWPIZ



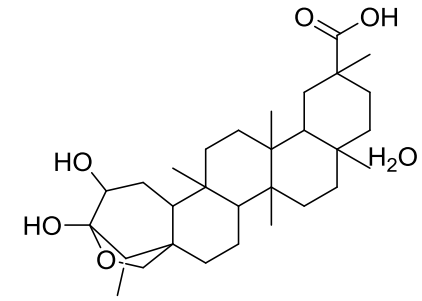
CAYHEQ



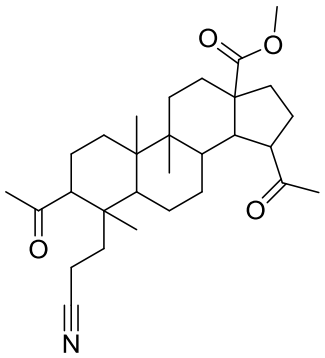
CEDWOZ



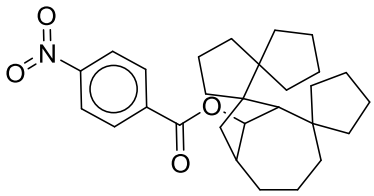
CERCEH



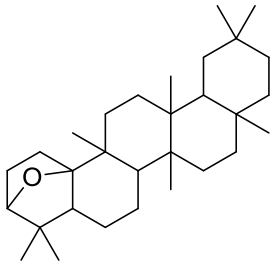
CEYVAD



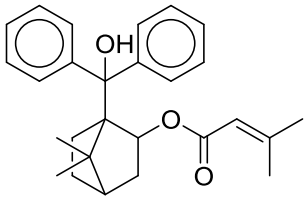
CICRIQ



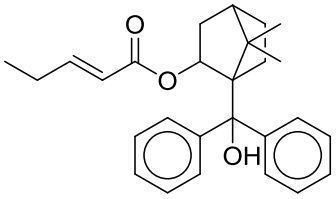
CIZZOB



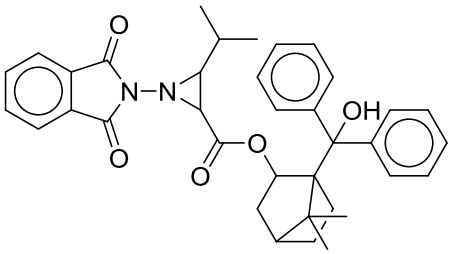
CMPANL01



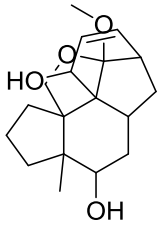
COBXOH



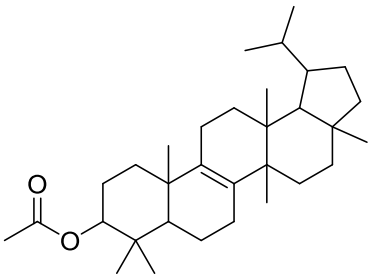
COBXUN

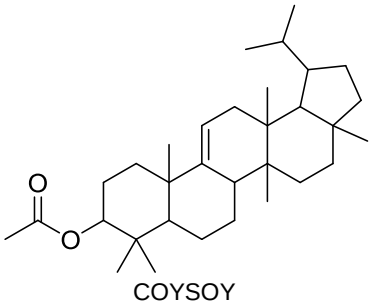


COBYAU

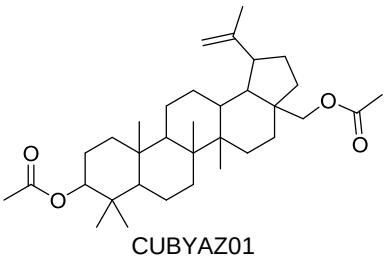


COCCAZ

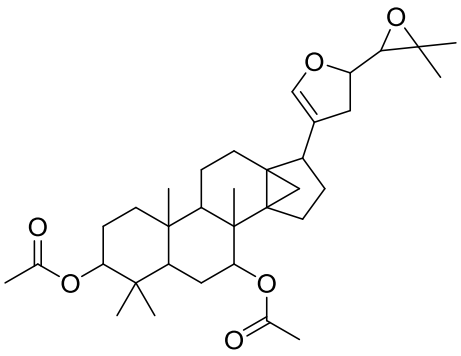




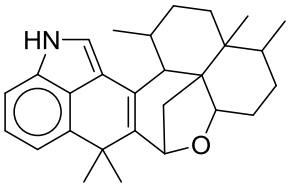
COYSOY



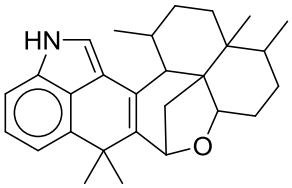
CUBYAZ01



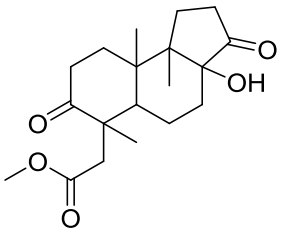
CULDAO



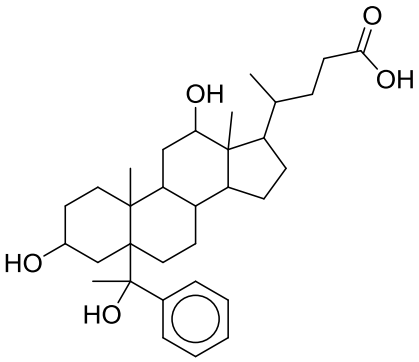
CUQMOS



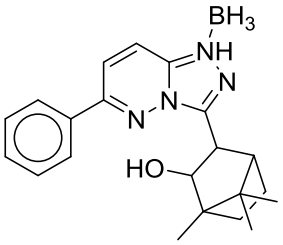
CUQNAF



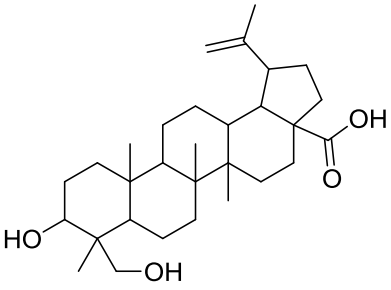
CUVPUE



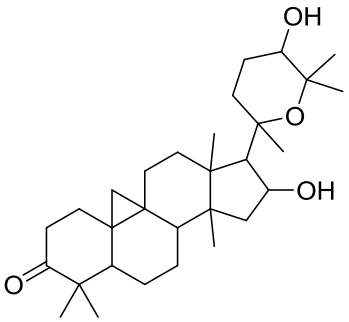
DAJLIJ



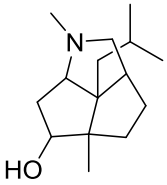
DAZPIE



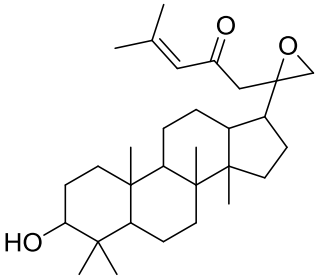
DENPUH



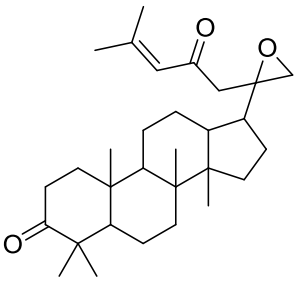
DEZDAN



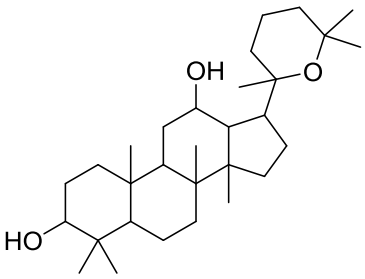
DIHVEV



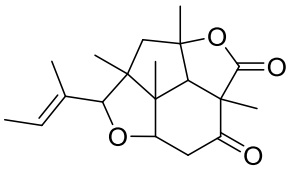
DIJWUP



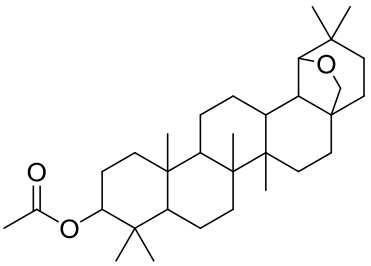
DIJXAW



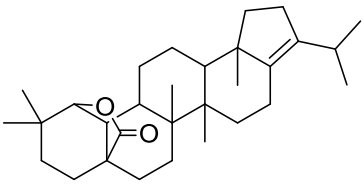
DOHDUZ



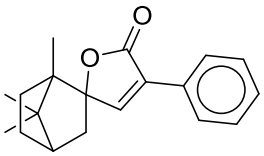
DORBUJ



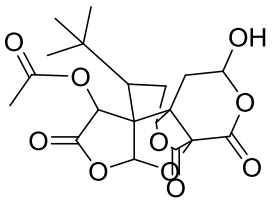
DOVGAX



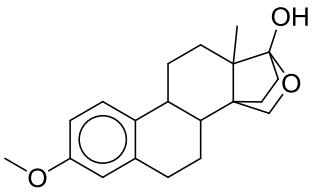
DOVGEB



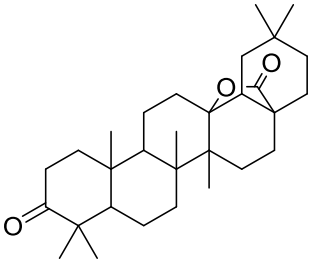
DUDKUJ



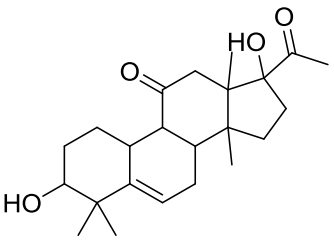
DUKKAV



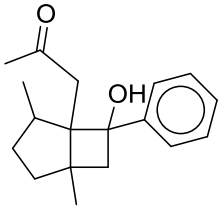
DULWAI



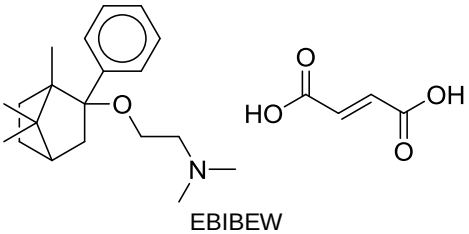
DUTXUM



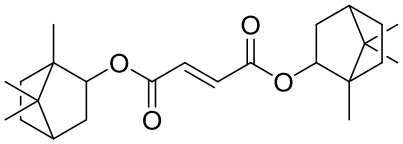
DXTMPO



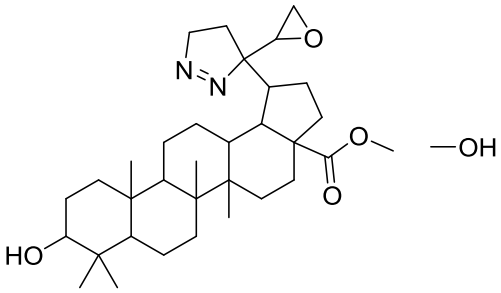
EADSUX



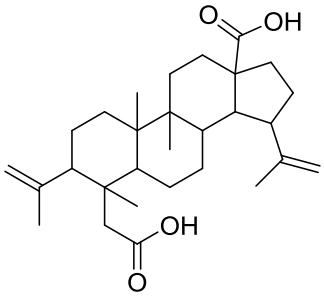
EBIBEW



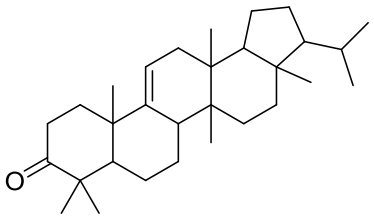
ECADIW01



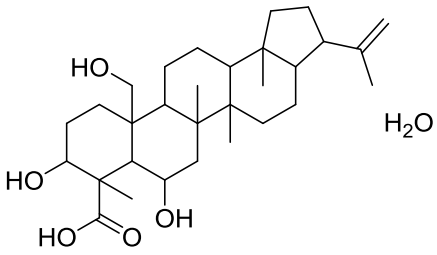
EDEGOK



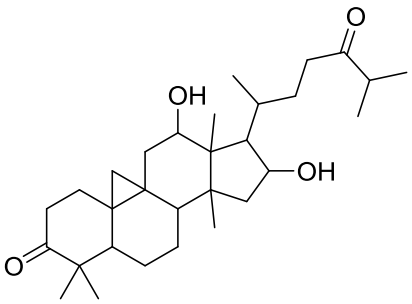
EDIFAZ



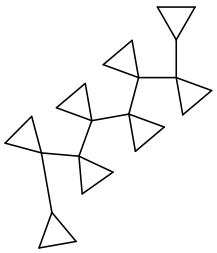
EFIGEG



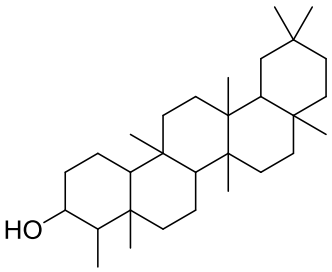
EGUFOD



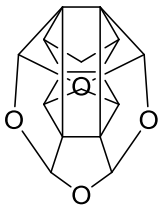
EGUFUJ



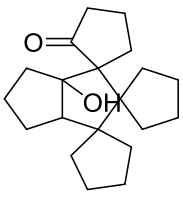
EMESIZ



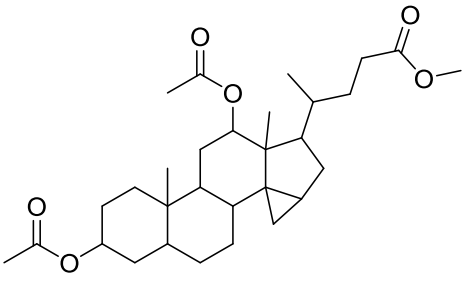
EPFRED02

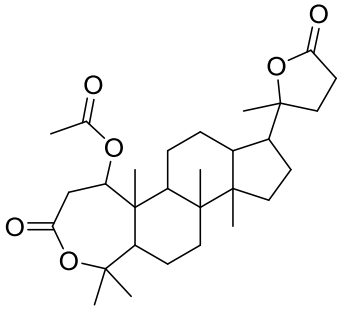


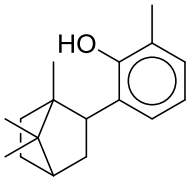
FAFJUR

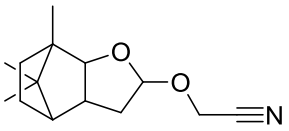


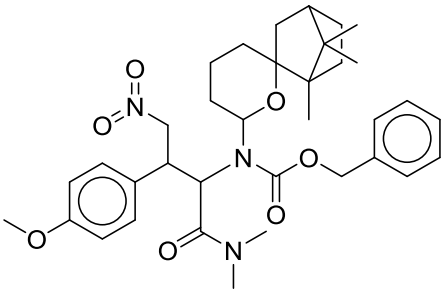
FASDAE

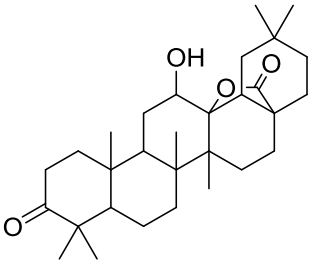

FATRUN

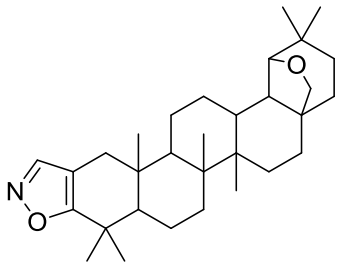

FAWLEU

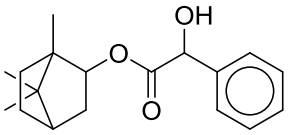

FEVCAL

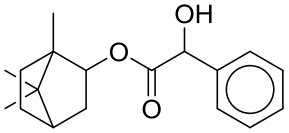

FIGLOW

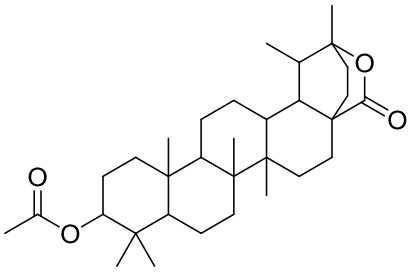

FIQTUV

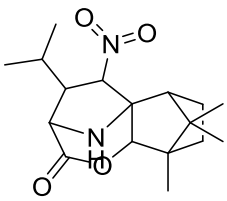

FITVOT

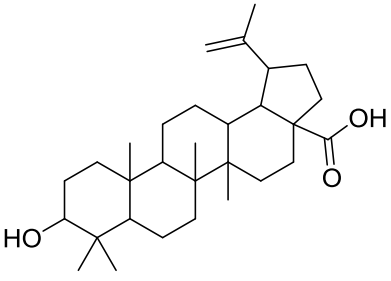

FOBQOC

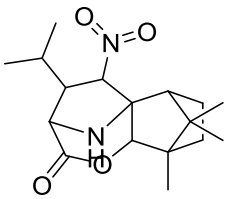

FOGHAL


FOHRAW

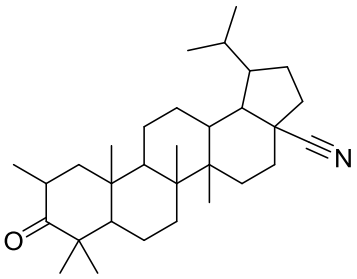

FOLVUX


FONYUD

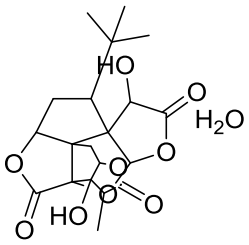

—OH
FOQKUT



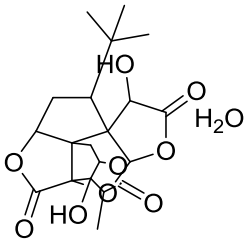
FOXCUR



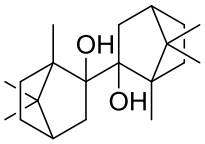
FUFGAO



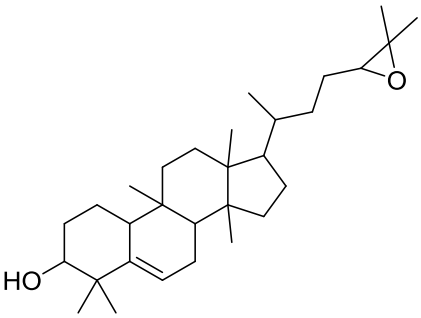
FUGTOQ



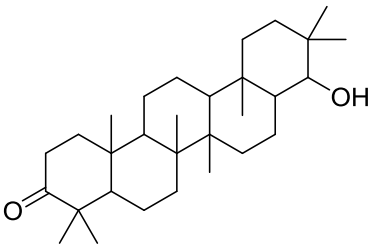
FUGTOQ01



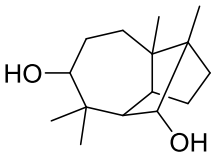
FUTZOJ



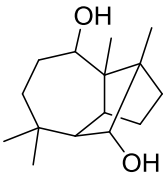
FUVVAT



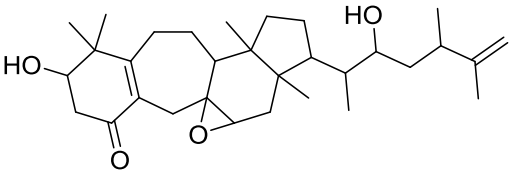
FUYNUI



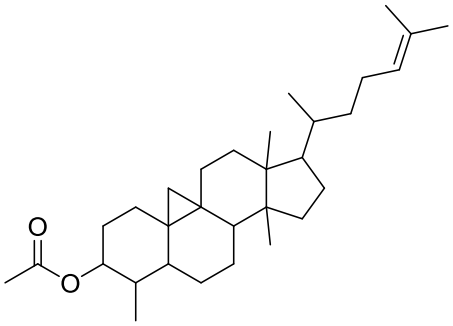
GACLII



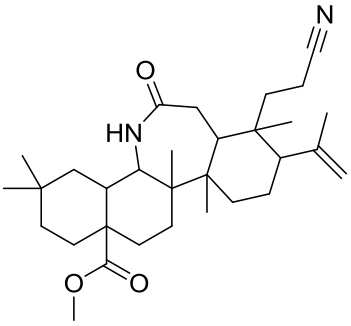
GACLOO



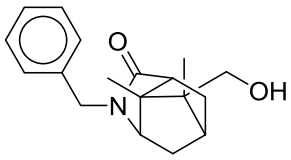
GACLUU



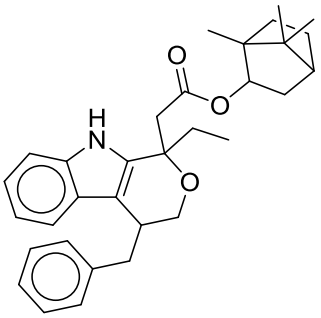
GAGQAG



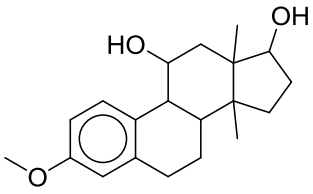
GARGAJ



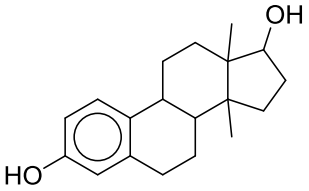
GATVAY



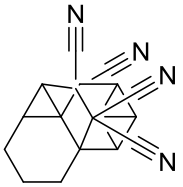
GAYCOY



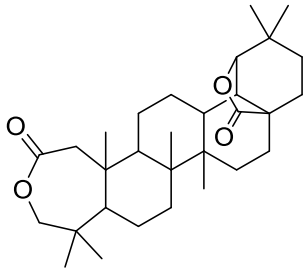
GAYKAS



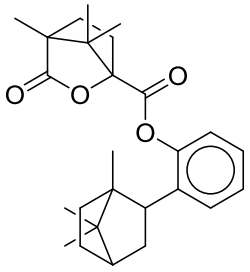
GAYKEW



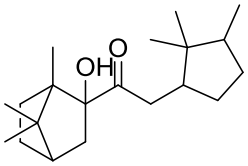
GEFMOT



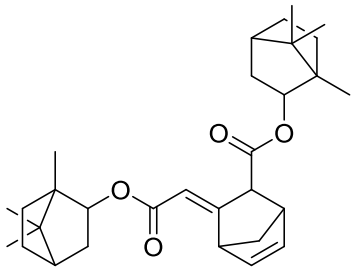
GEJLAK



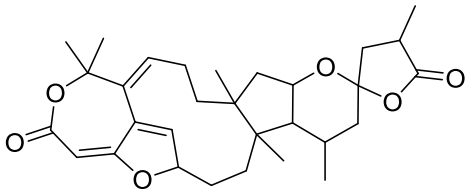
GEKRAR



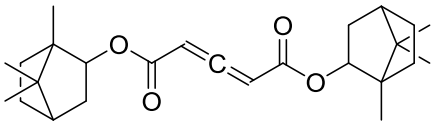
GETBEM



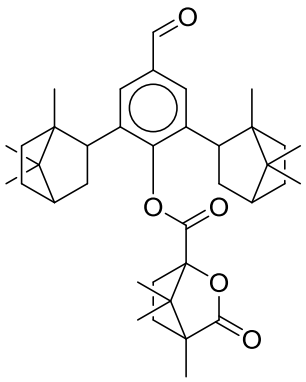
GETCUE



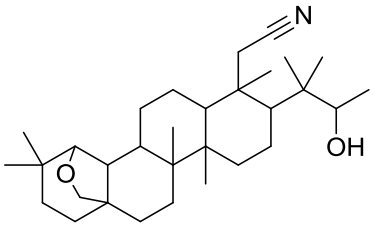
HADJEC



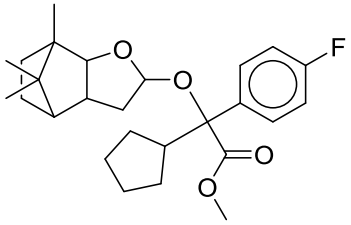
GEVNAX



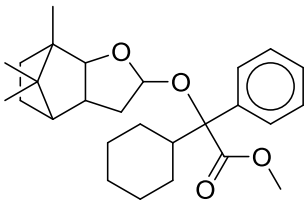
GIMJAP



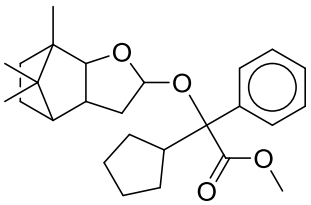
GIXNIM



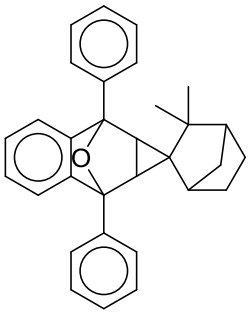
GOGPUP



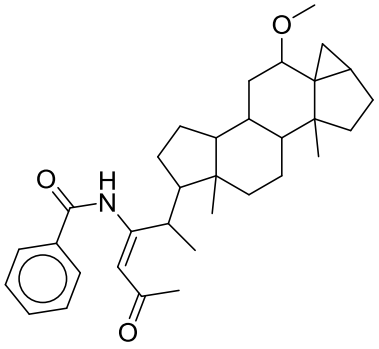
GOGQAW



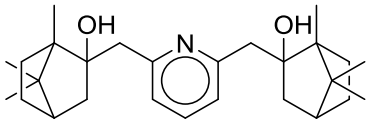
GOGQEA



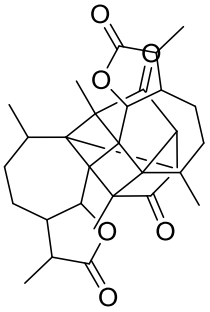
GOKNEZ



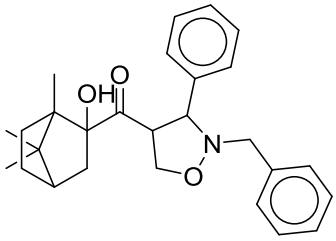
GOKXIN



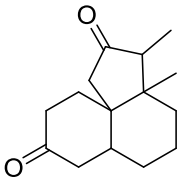
GUXGUB



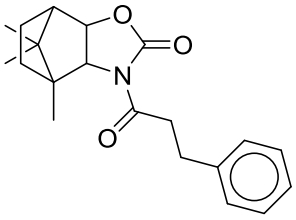
GUYFIR



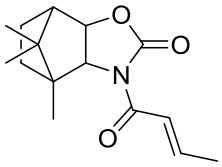
HAVVUW



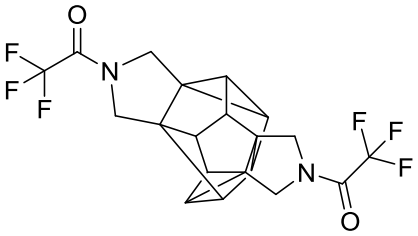
HAVZEJ



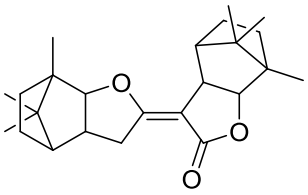
HERRUR



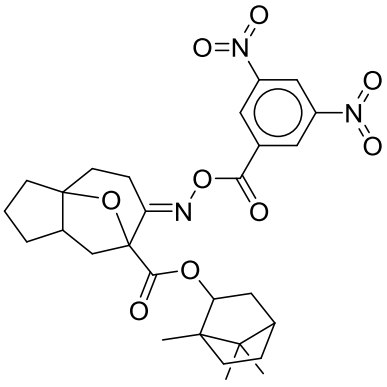
HERSAY



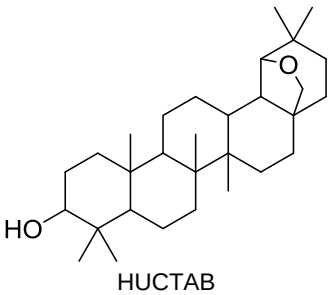
HEXSAF



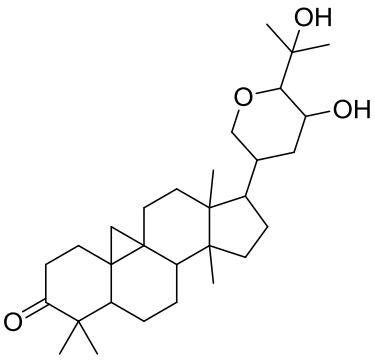
HIFCII



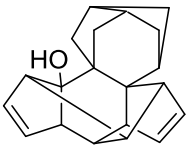
HOTJAB



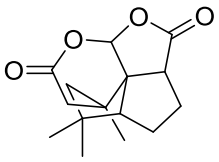
HUCTAB



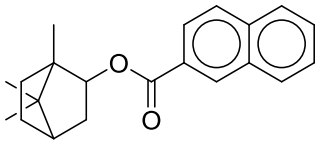
HULXIX



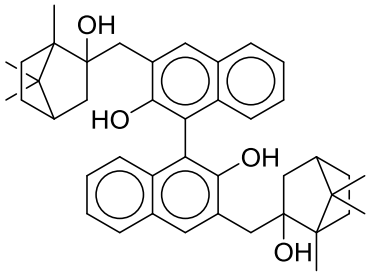
HUXQAU



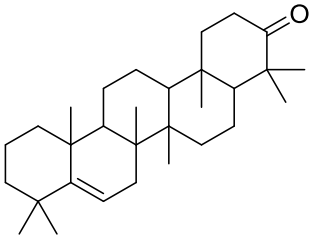
IFAHAZ



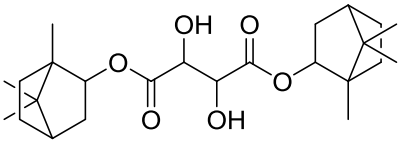
ISELEX



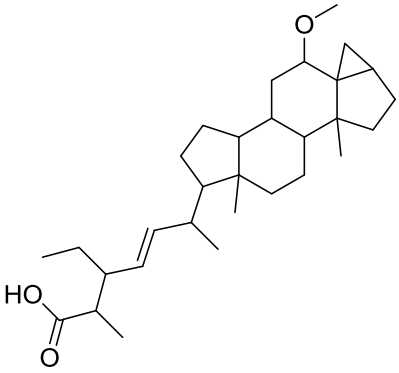
JAGQOX



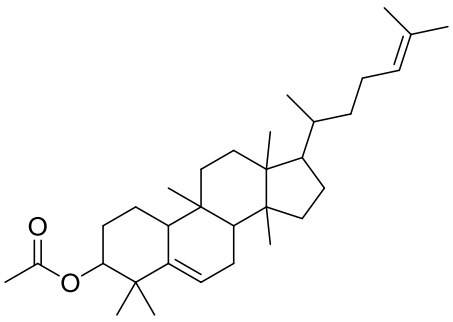
JASSOL



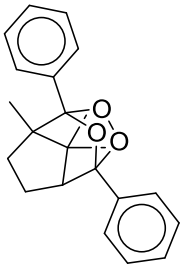
JASWUV



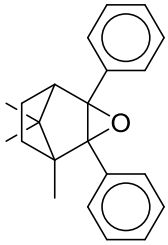
JAYVOU



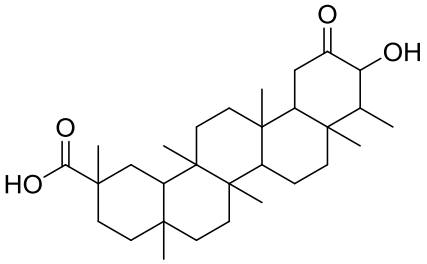
JAZTIN



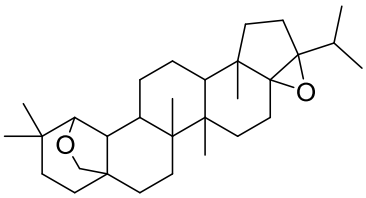
JESGIX



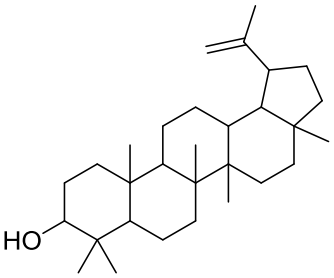
JESGOD



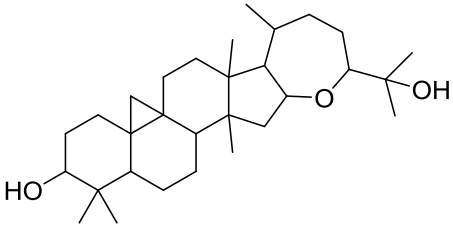
JIBBOL



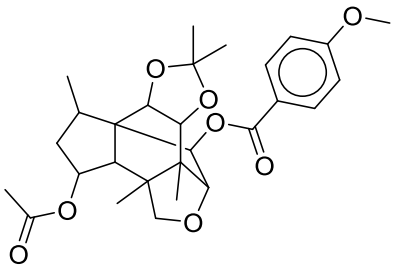
JIRKOL



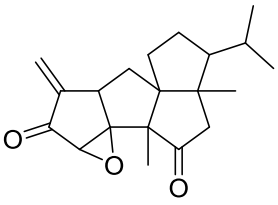
JOLBIW



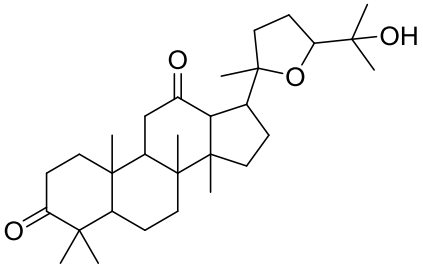
JOQPIQ



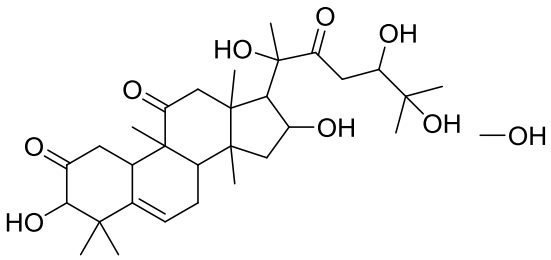
JUFQAE



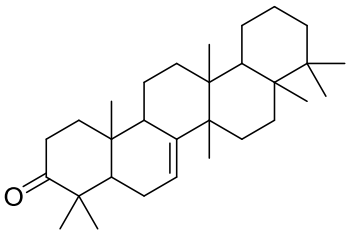
JUVFAH01



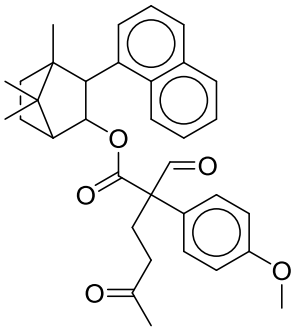
JUWDOU



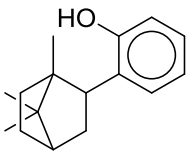
KAGNAJ



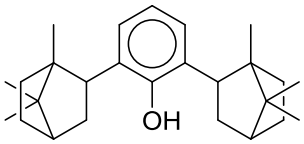
KAGZUN



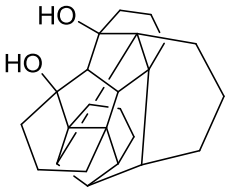
KANDUY



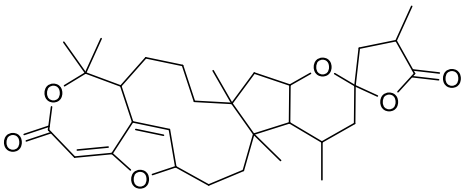
KANJOZ



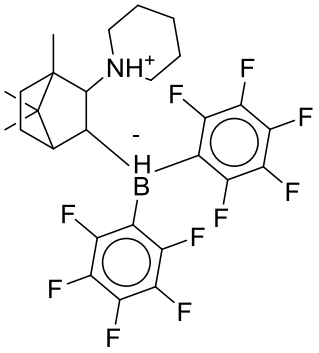
KANJUF



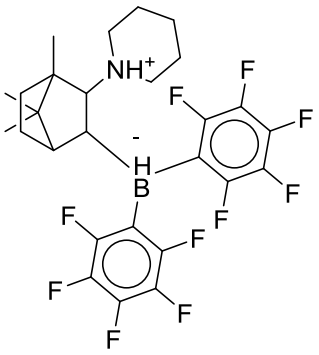
KEPSON



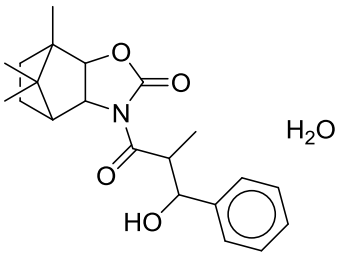
KEWREJ



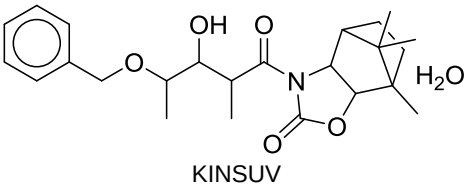
KICDEH



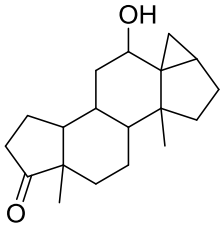
KICDIL



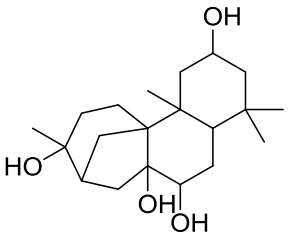
KINSEF



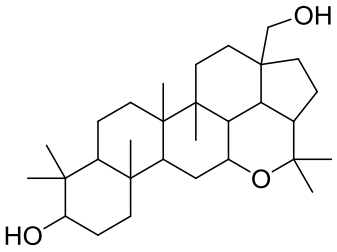
KINSUV



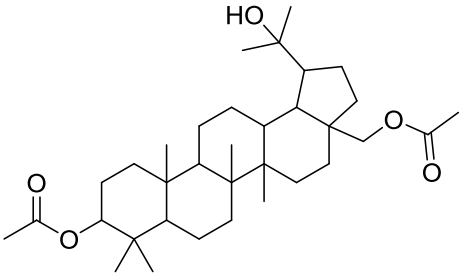
KONWAL



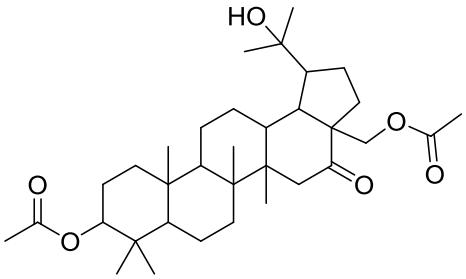
KOPKUV



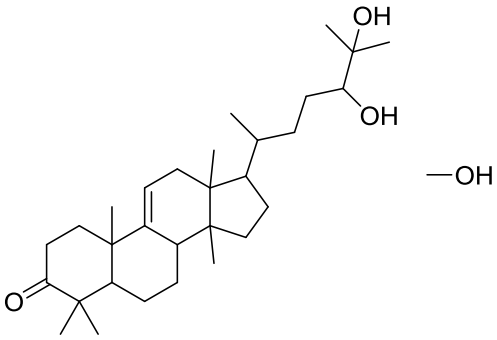
KOWBUV



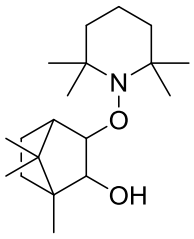
KOWCAC



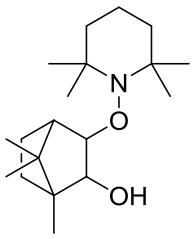
KOWDAD



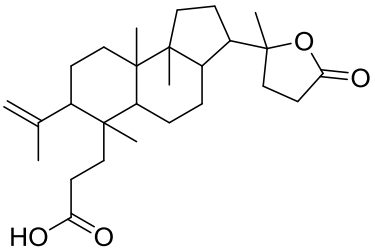
KUKJOR



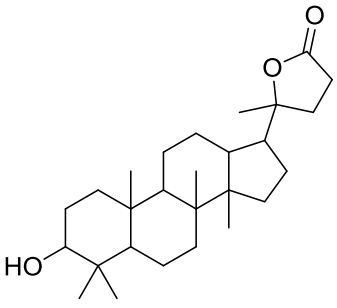
KUMTAP



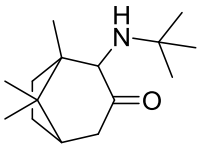
KUMTIX



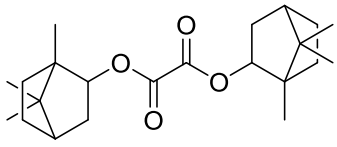
KUTMIW



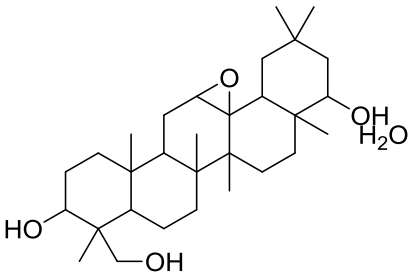
KUZKAS



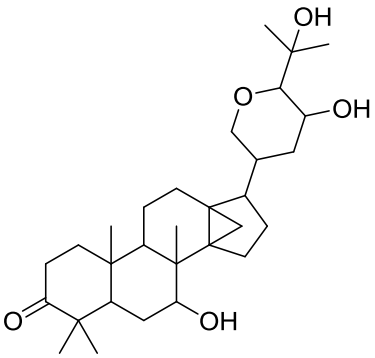
LADNIP



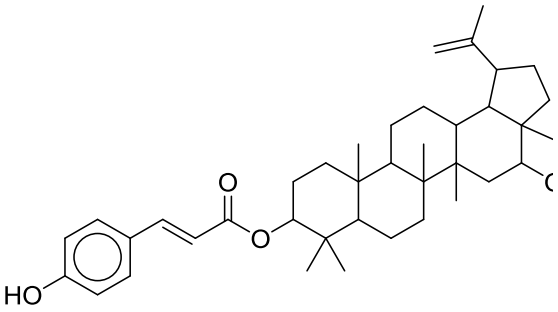
LAKGIO



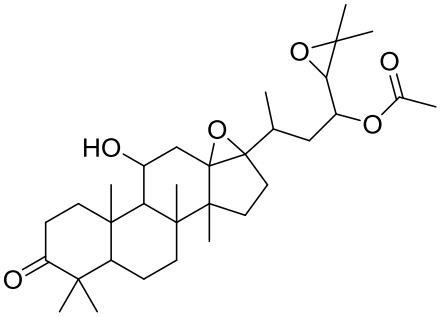
LAQYUY



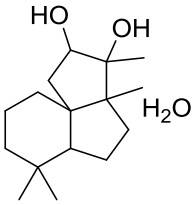
LEBWEV



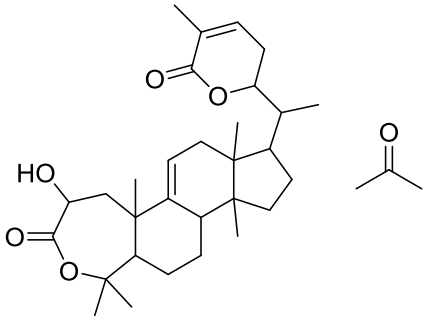
LEFHAG



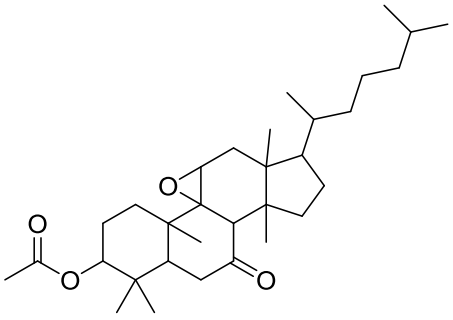
LEPJAS10



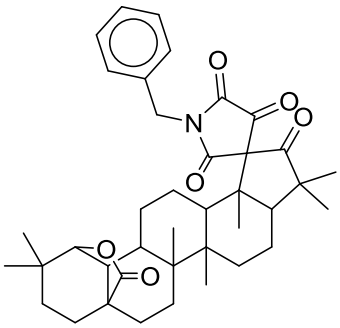
LEPYUA



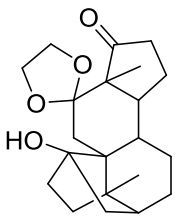
LIBMIU



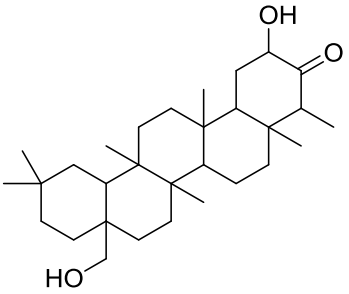
LICKAK



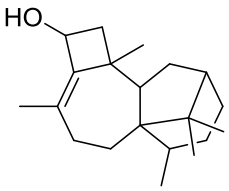
LIGCAH



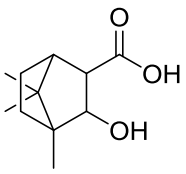
LITQIP



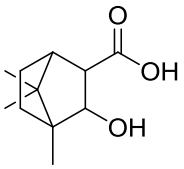
LIXMIP



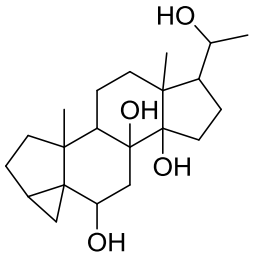
LOLPAF



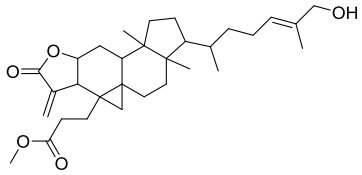
LONPEJ



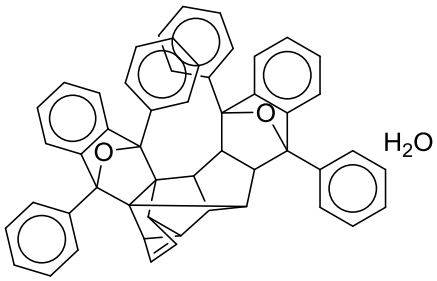
LONPIN



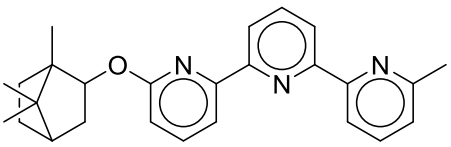
LUDTUA



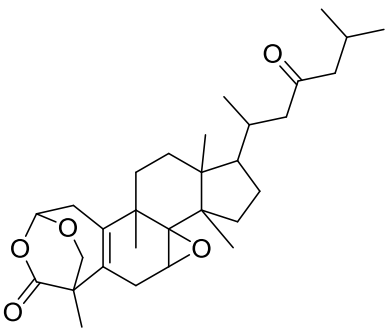
LUDWAJ



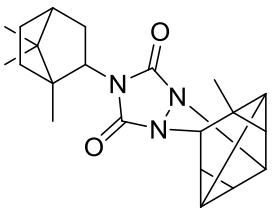
MANGOY



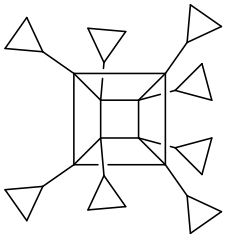
MAXJUQ



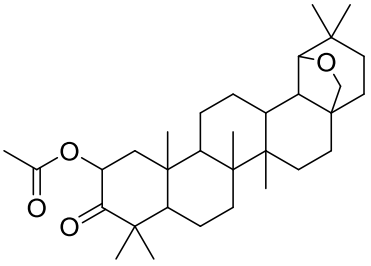
MAZCIB

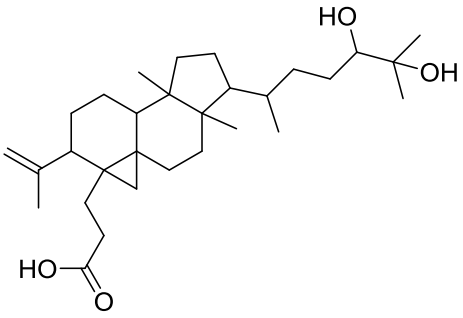


MBLVBT



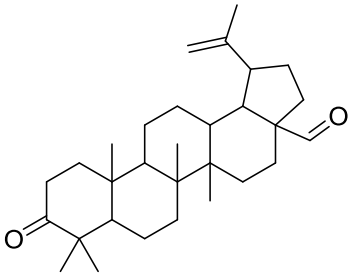
MIDYAA



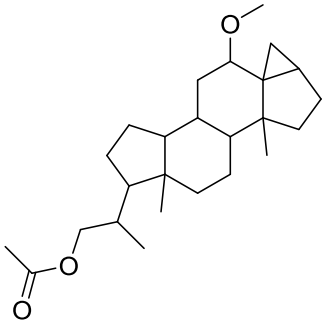


MOFYOX

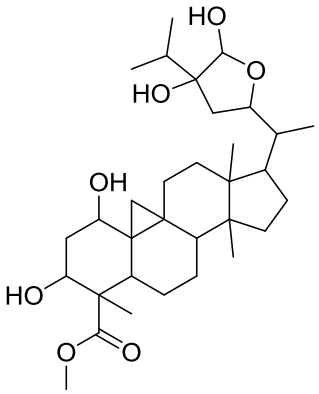
MIRMOQ



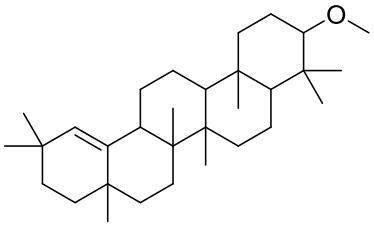
MOHZOA



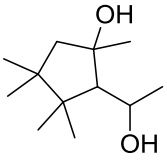
MOZLIX



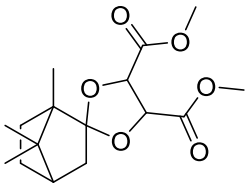
MPASSF10



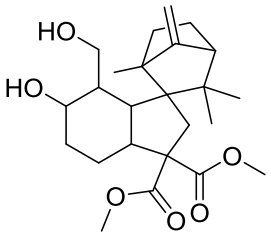
MUFXUG



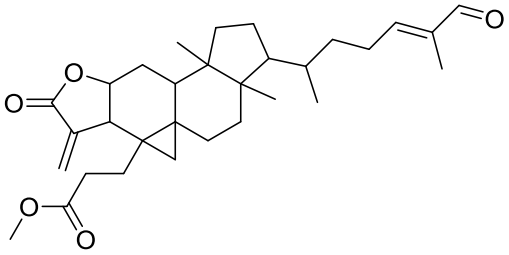
NABHUU



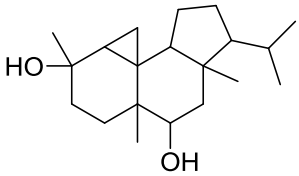
NAFWEW



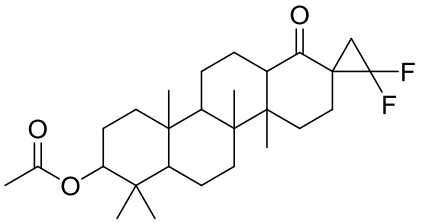
NAHYOK



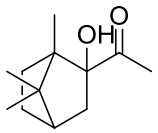
NAMVUS



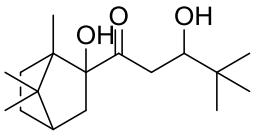
NEMXUY



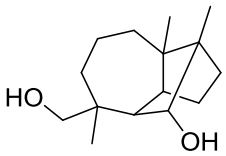
NEXHUV



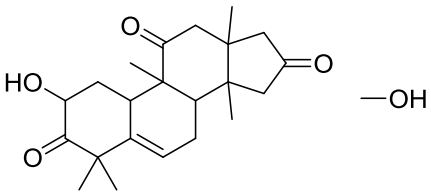
NEXNIN



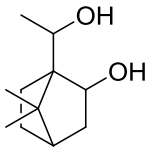
NEXNOT



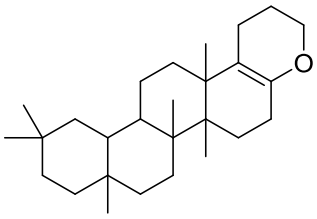
NEZFON



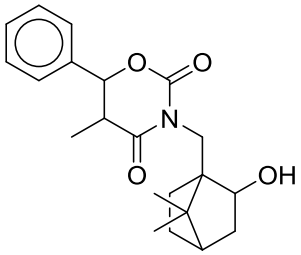
NIWZAV



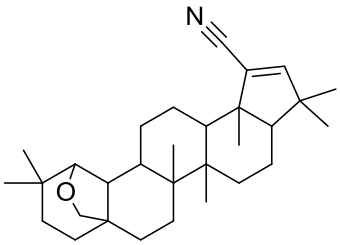
NOBYEK



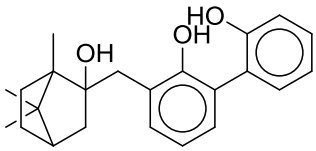
NOKVUE



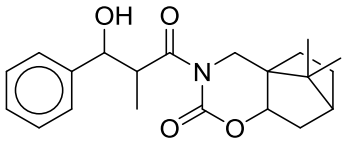
NOKWEP01



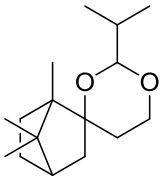
NUCXOA



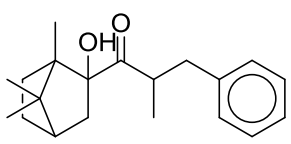
NURXED



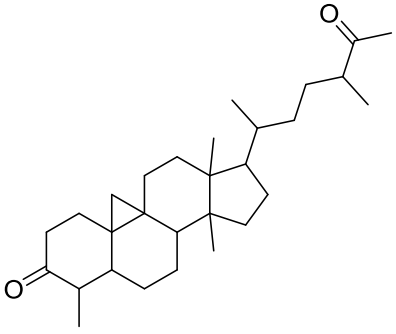
OCAXUM

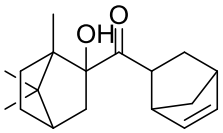


OCILER



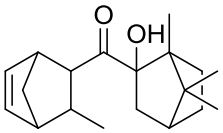
OCUWAK



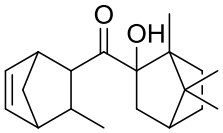


OGITIH

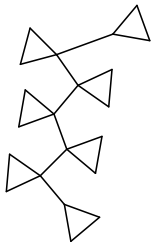
OFOQIL



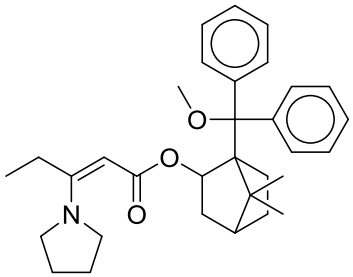
OGITUT



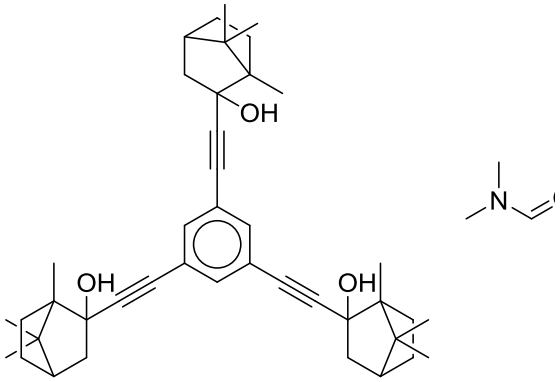
OGIVAB



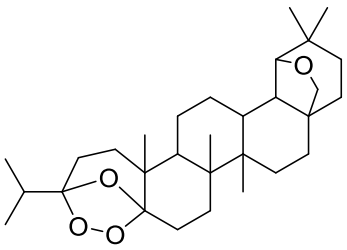
OJIXIP



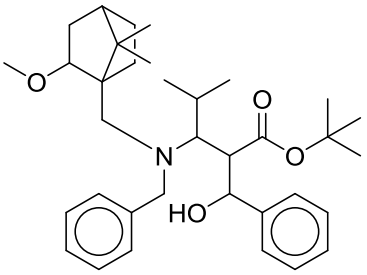
OKAJIU



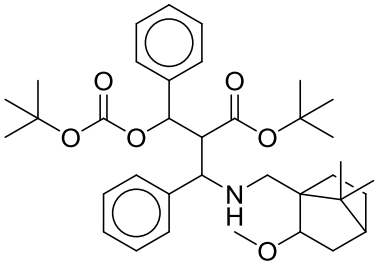
OKITIL



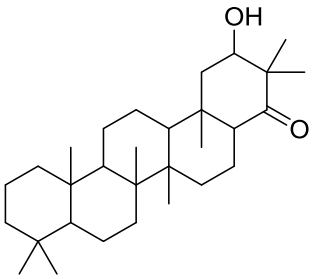
OLESUU



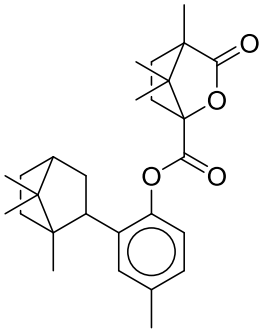
OMULIS



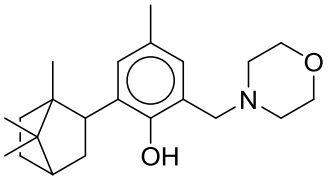
OMULUE



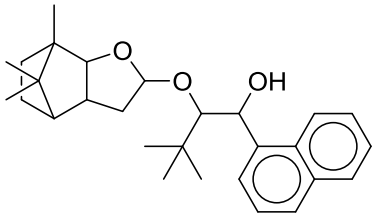
OPELEB



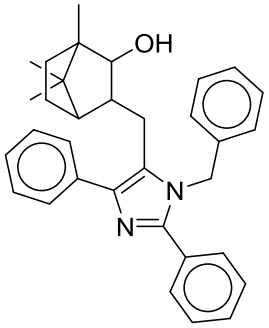
OSUVEE



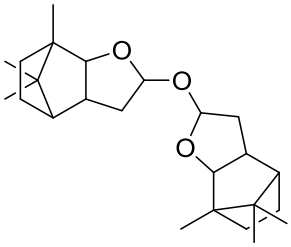
OSUVOO



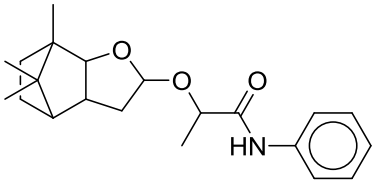
PABBEA



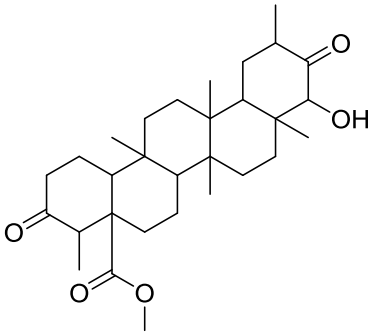
PABBIE



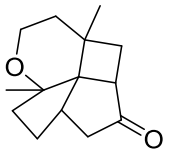
PABDAY



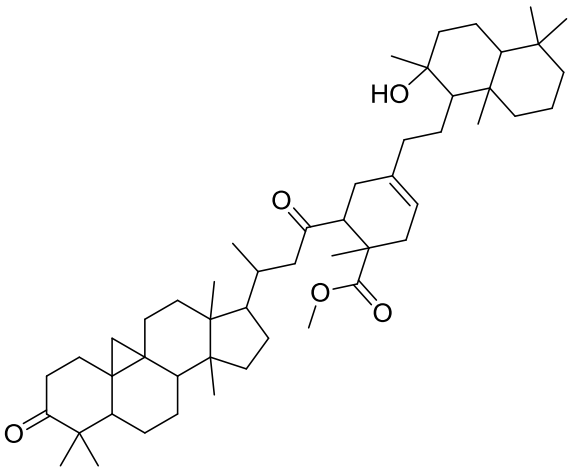
PABFII



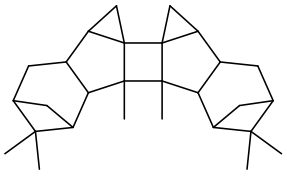
PAPGAO



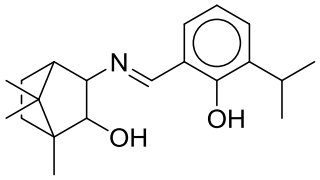
PEFCAE



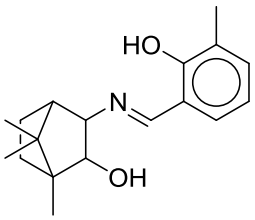
PENCAO



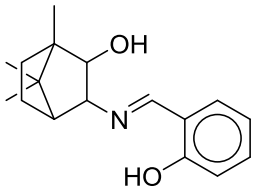
PIMPAE



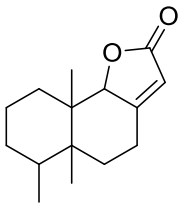
PIZFIP



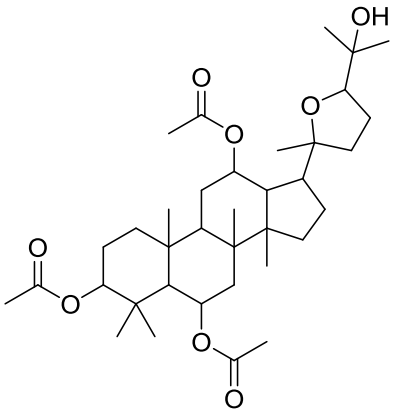
PIZFOV



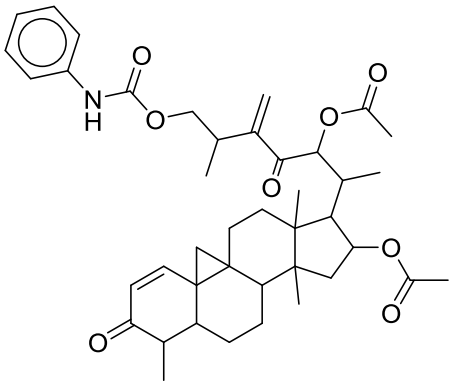
PIZFUB



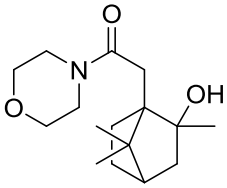
POPZAX



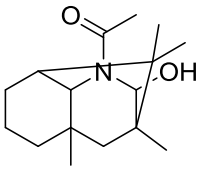
QAJSIF



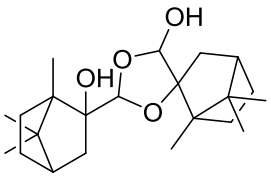
QAQHIB



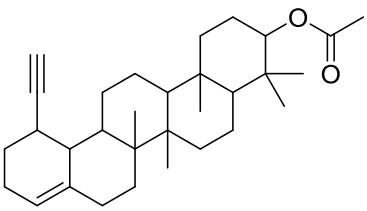
QEMQUU



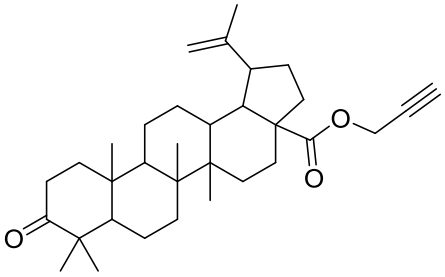
QIQCIC



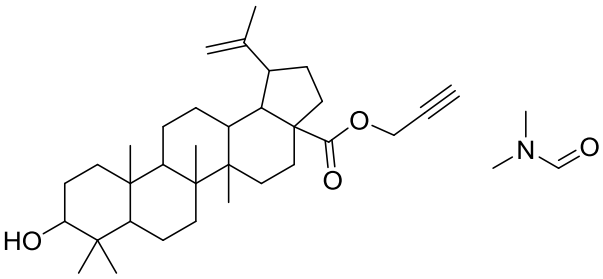
QUMLEQ



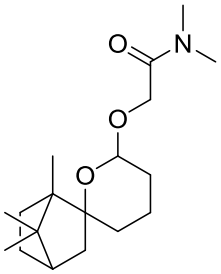
QUQYAD



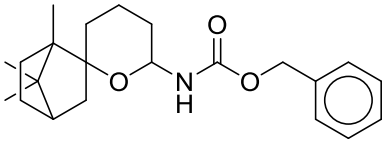
QUYTEL



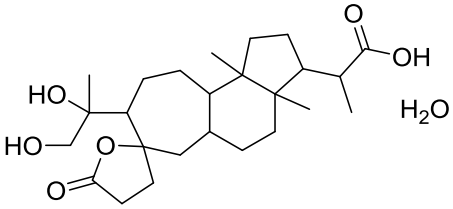
QUYTIP



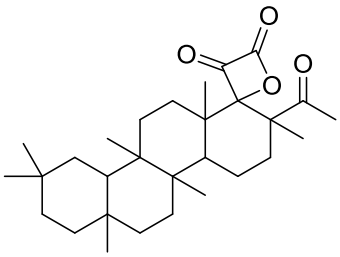
RALGER



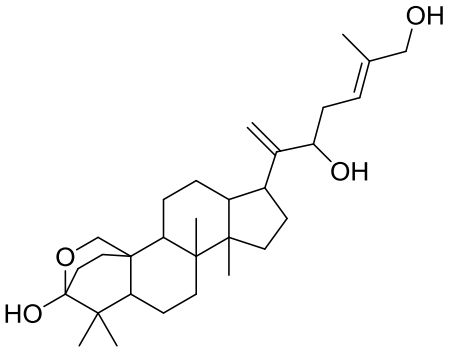
RALGIV



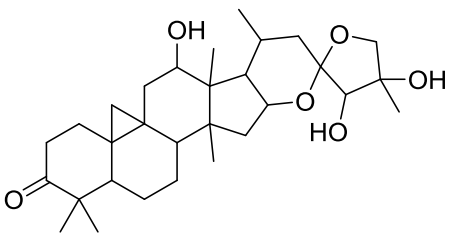
RAMGOC



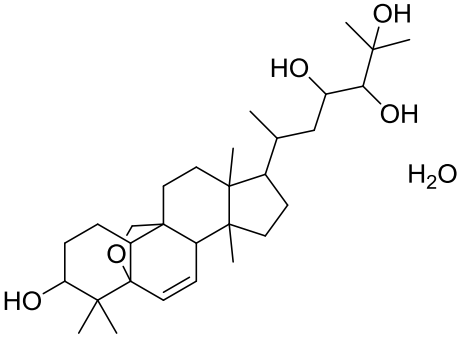
REXMEO

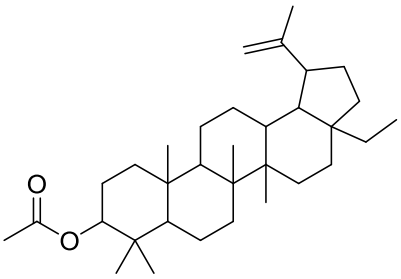


RHSLCT10



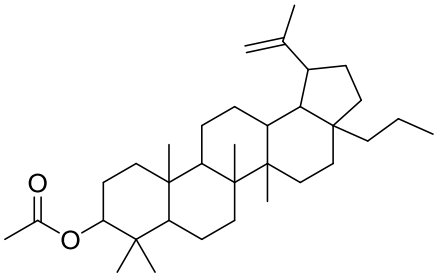
RIMGUR



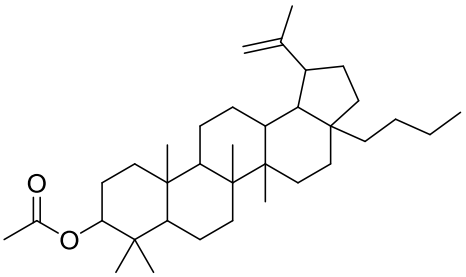


ROJSIU

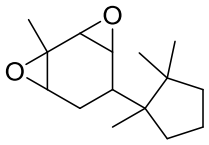
RIWCIK



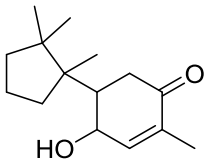
ROJSOA



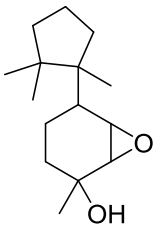
ROJSUG



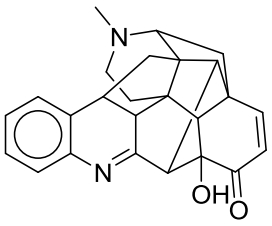
RULGOU



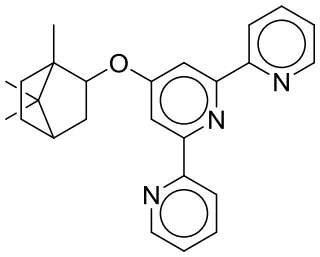
RULGUA



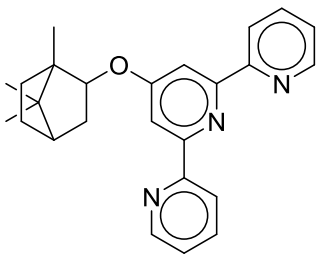
RULHIP



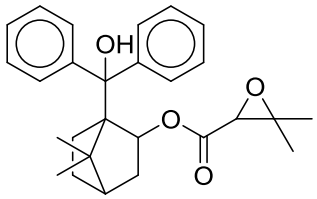
RULNOB



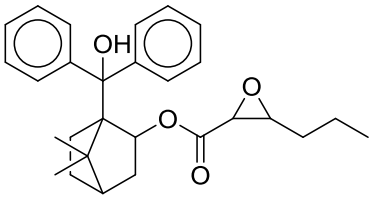
RUSYOT



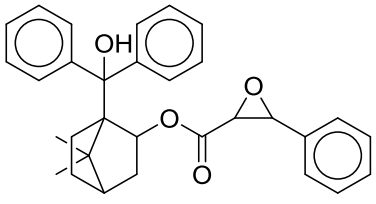
RUSYUZ



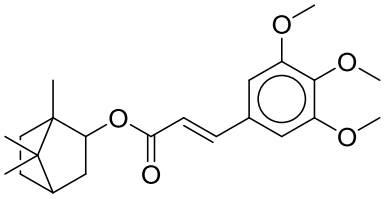
SAFBEH



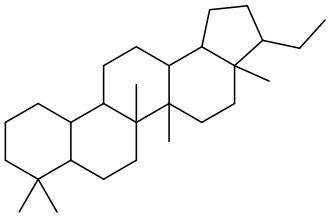
SAFBIL



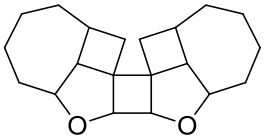
SAFBOR



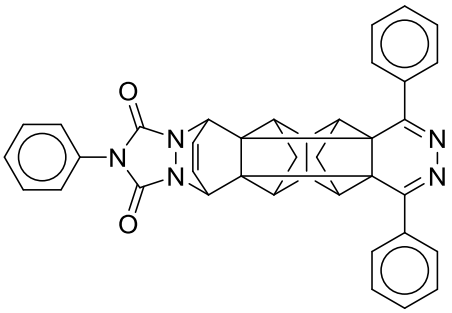
SASLEF



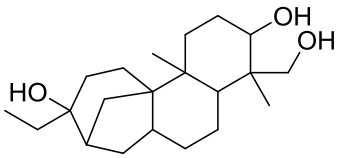
SEKZUD



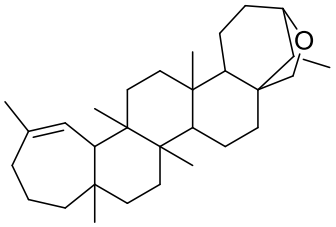
SEMKOK



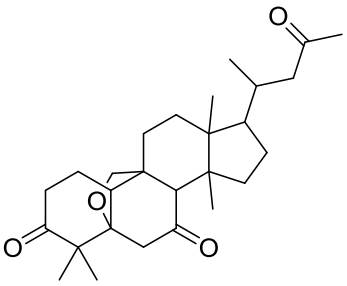
SIFFET



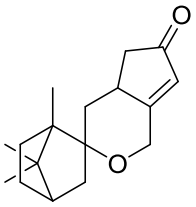
SIQGUU



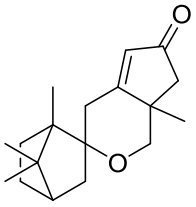
SLSPRM



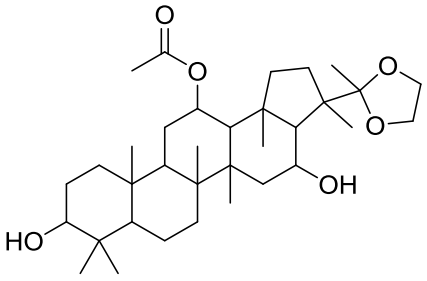
SOQKOZ



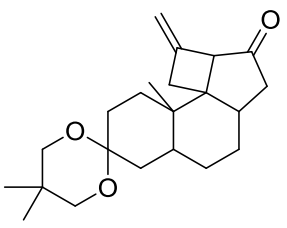
SORROH



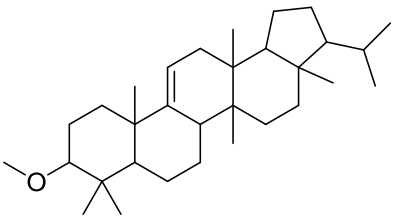
SORRUN



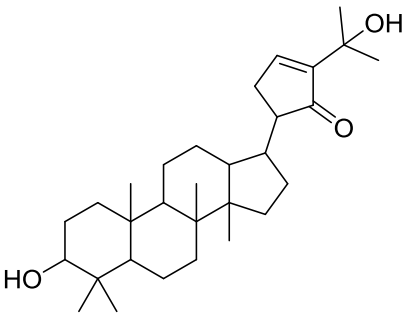
SPEREK10



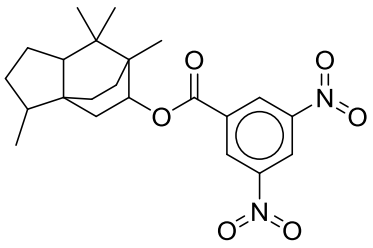
STMDIN



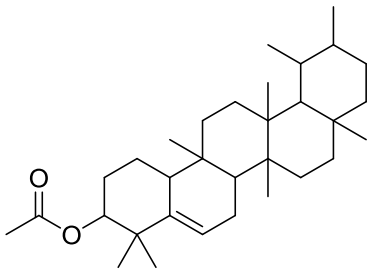
SUZBOF



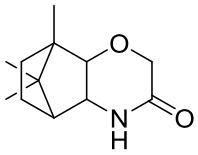
TAGTOK



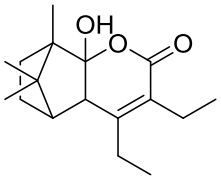
TAYTIY



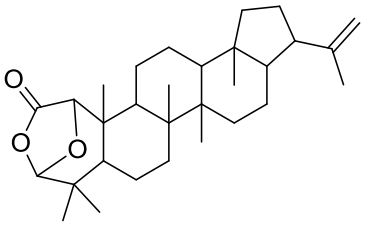
TEGZOU



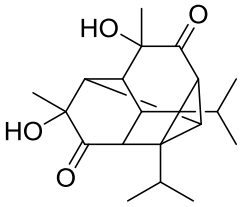
TEMKAZ



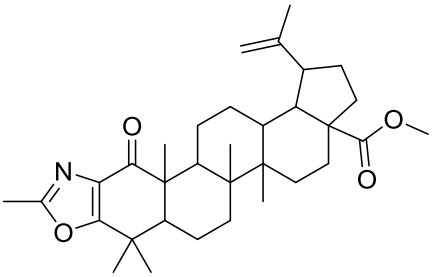
TESLOT



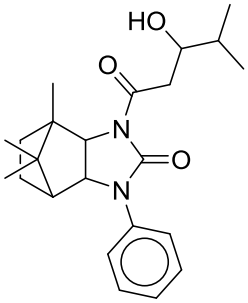
THYSNL



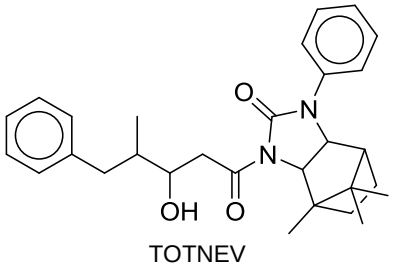
TIFNUS



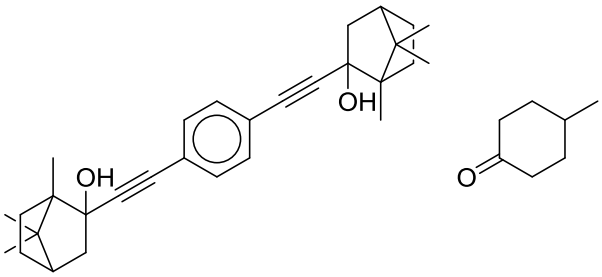
TOQBUY



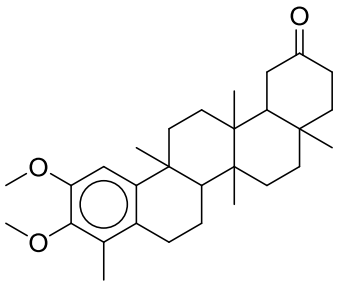
TOTNAR



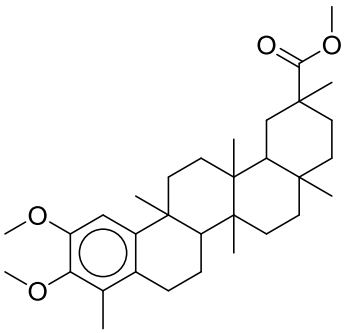
TOTNEV



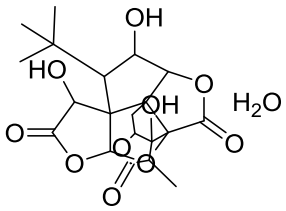
TUHWAW



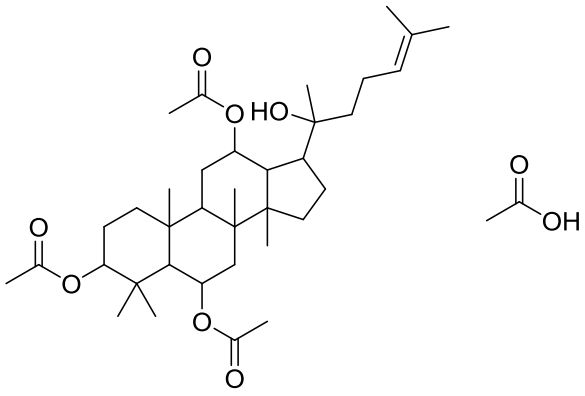
TUQZEM



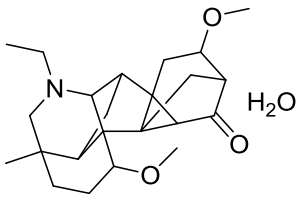
TUQZOW



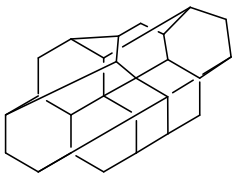
UFEHIW



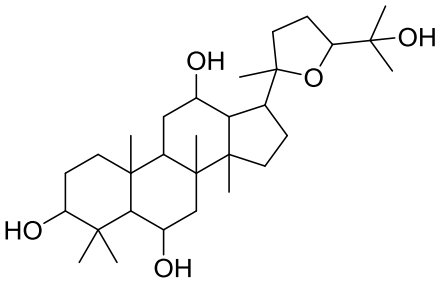
UFUCIJ



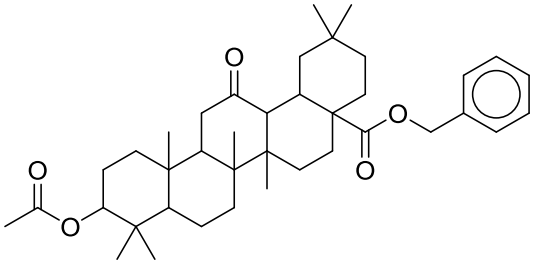
UFUQAO



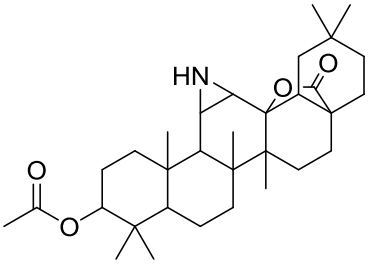
UJISAH



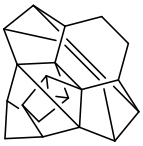
UPAXOZ



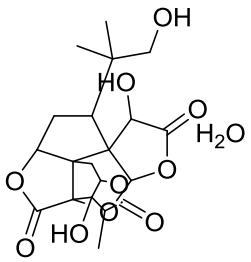
URACOG



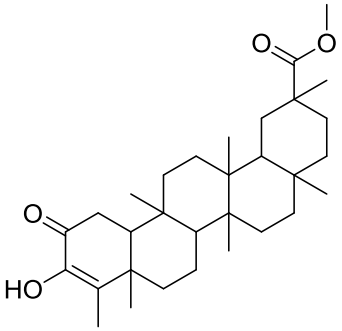
URENOV



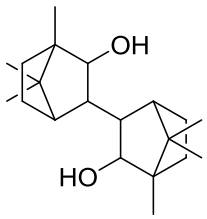
VABNER



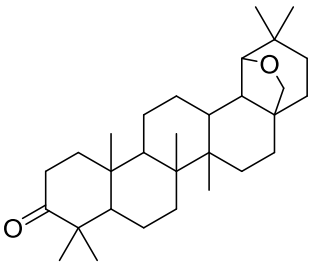
VATGEE



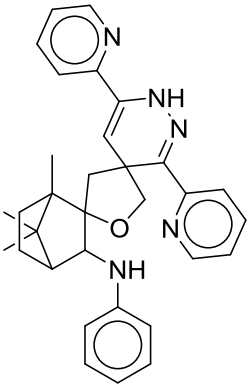
VEFNOJ



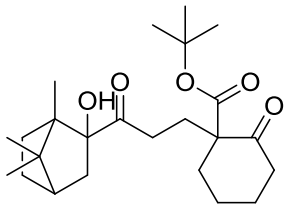
VEKMED



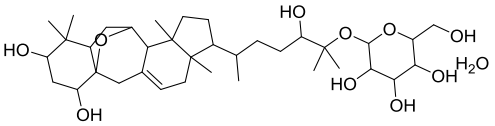
VEPBEX



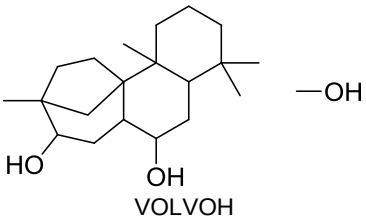
VIWGIS



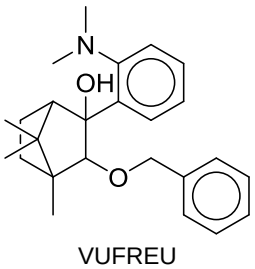
VOHFUU



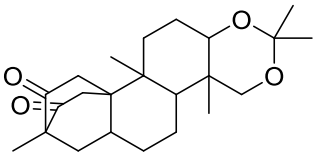
VOKZAX



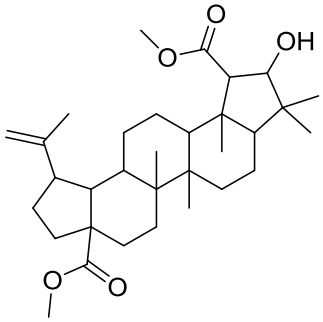
VOLVOH



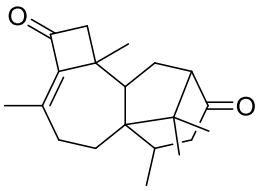
VUFREU



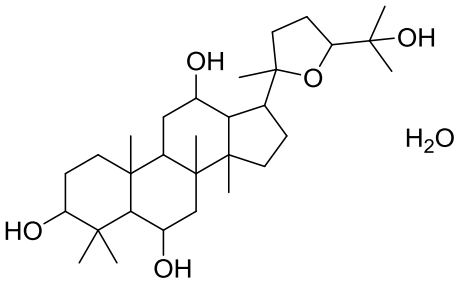
VUGWUP



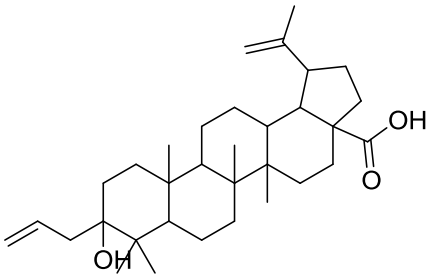
VUKCUZ



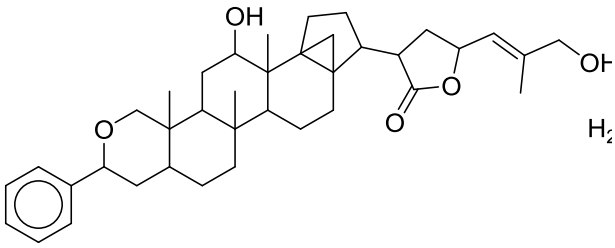
WABKAL



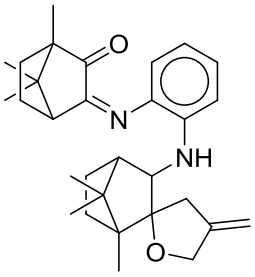
WADGIT



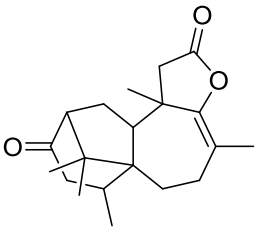
WAKYUE



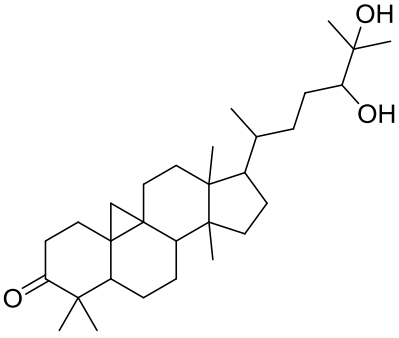
WINPUG



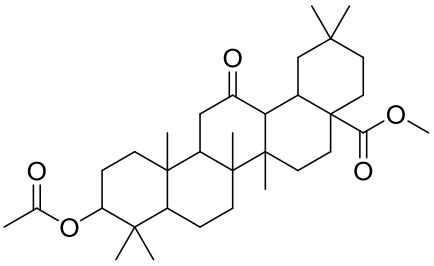
WIXWEG



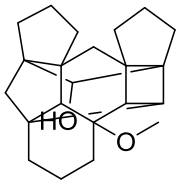
WIYKEW



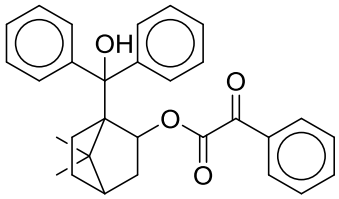
WOBKAZ



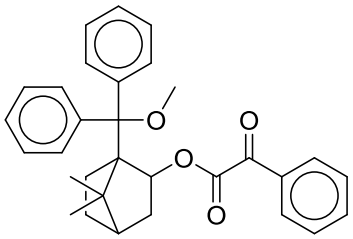
WOBKIH



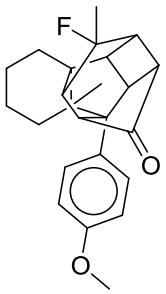
WOMSUM



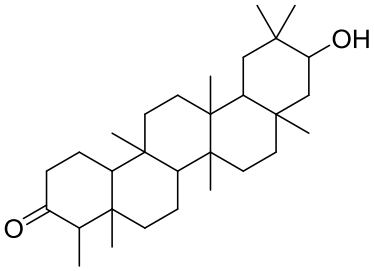
WORYIL



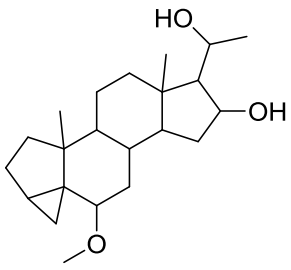
WORYOR



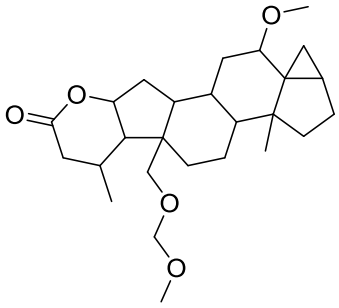
WOVFIY



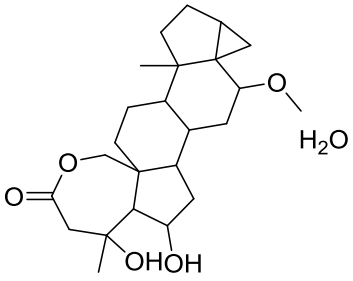
WOXKEZ



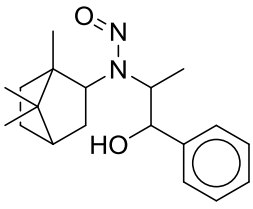
WUDKEN



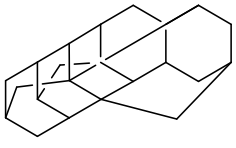
WUDKIR



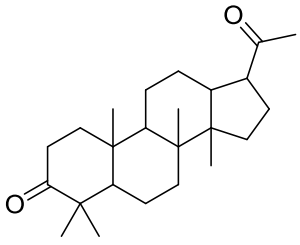
WUDKOX

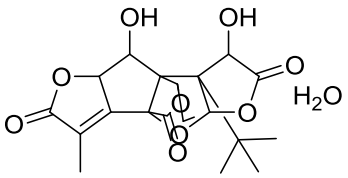


WUFMEP

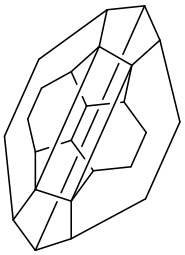


WULYEI

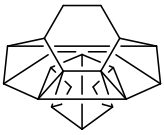




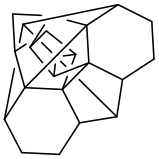
XAHTIM



WUTVUE

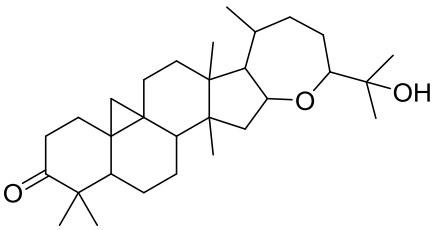


XAZXAX

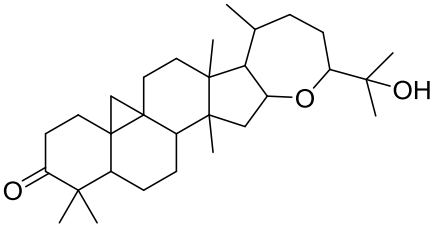


XAZWUQ

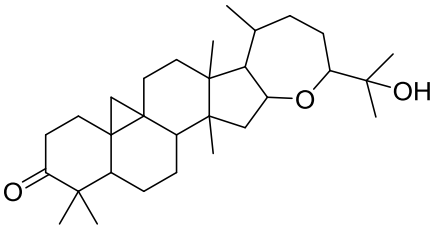
XAZXEB



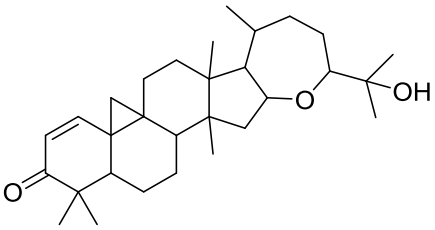
XEBMIB



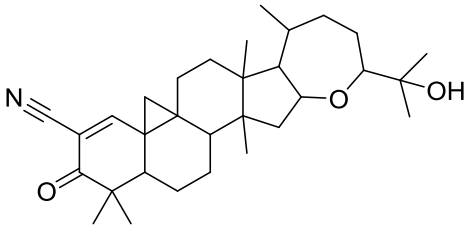
XEBMIB01



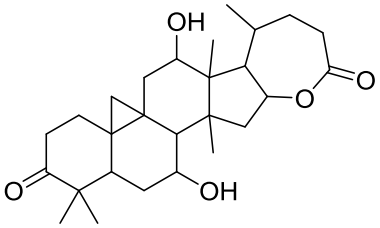
XEBMIB02



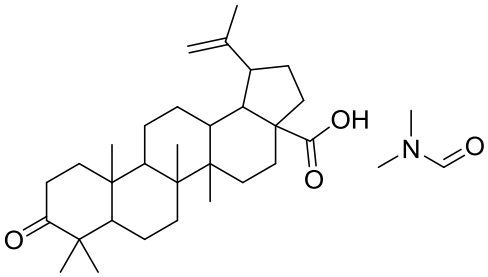
XEBMUN



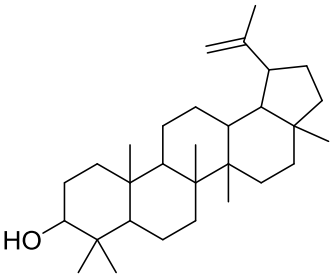
XEBNIC



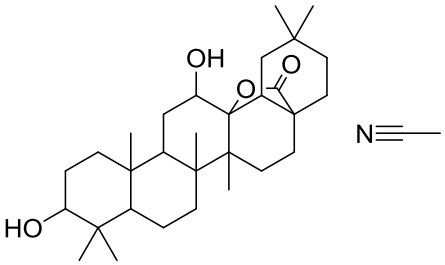
XEDKUN



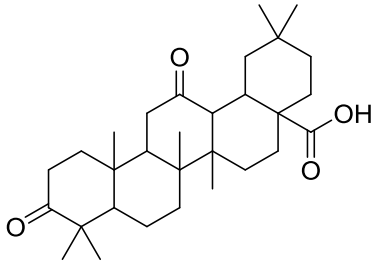
XEJNUX



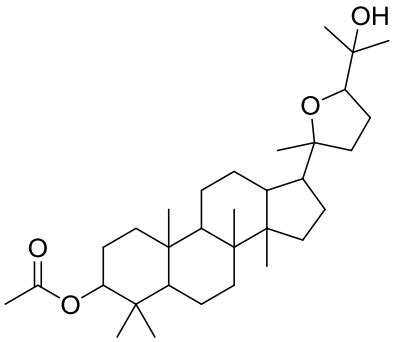
XENQOX



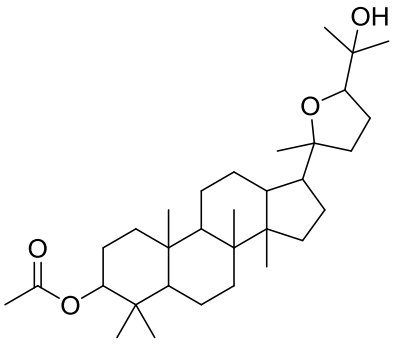
XEPGUW



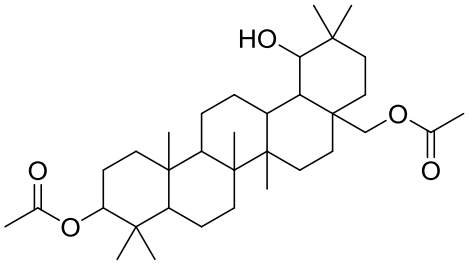
XEPHAD



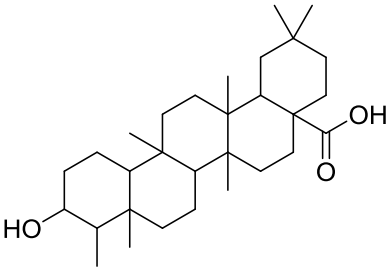
XEVJEN



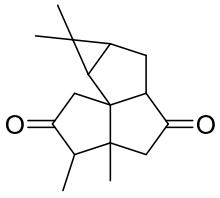
XEVJEN01



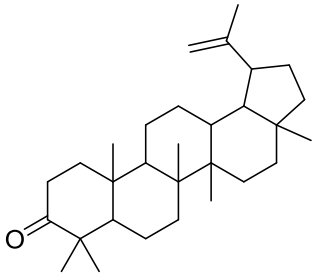
XICPAB

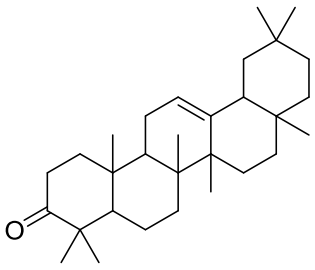


XIZDAM

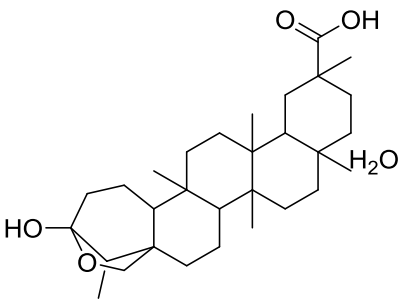


XONXON

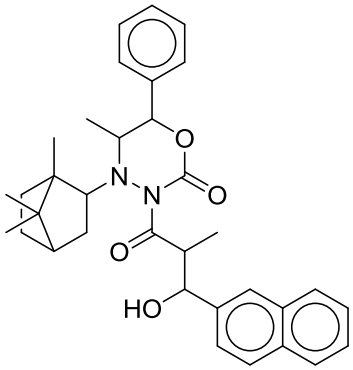




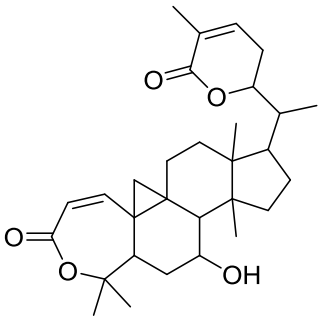
XUKNEY



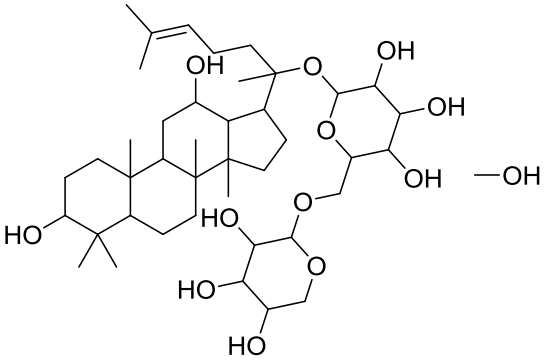
YACNEV



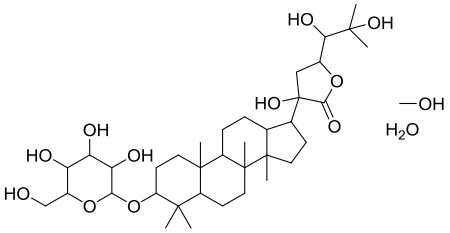
YAJZEP



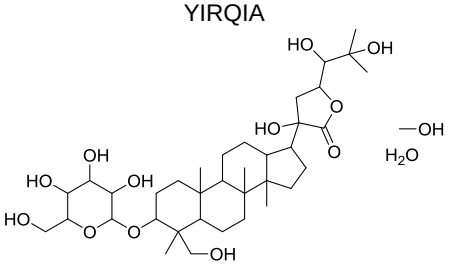
YIRLOC



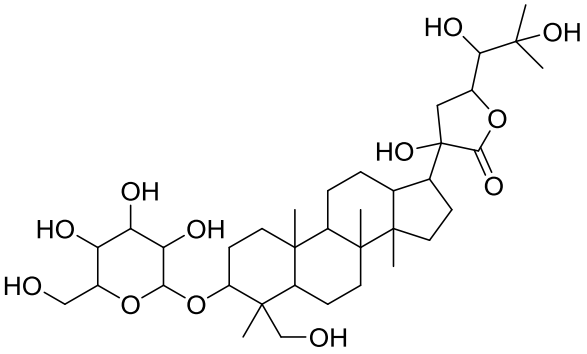
YIRQIA



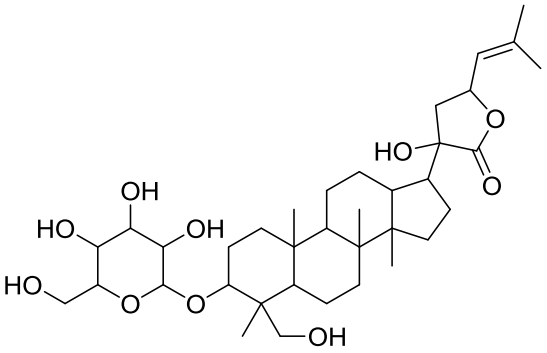
YOLLES



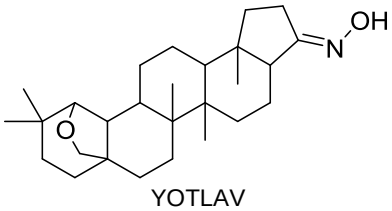
YOLLIW



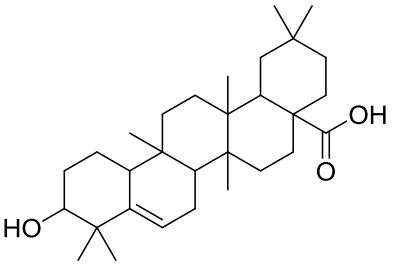
YOLLOC



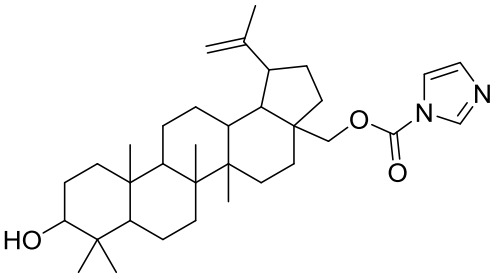
YONQUP



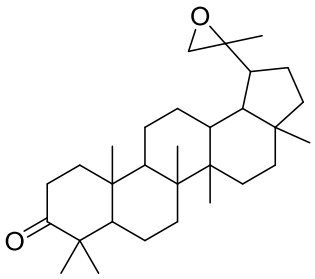
YOTLAV



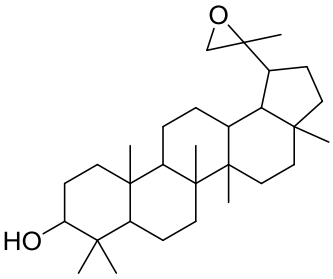
YOTPON



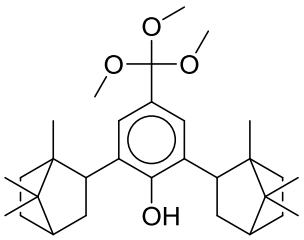
YUQRAE



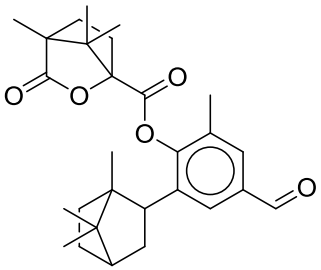
ZAYZIK



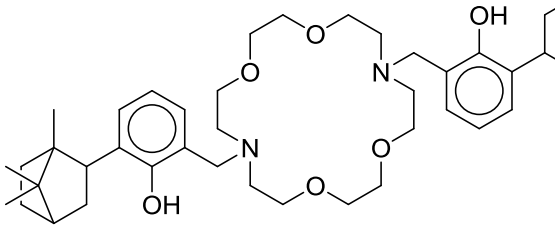
ZAYZOQ



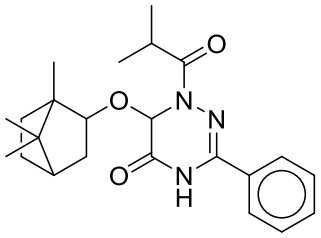
ZEPTOF



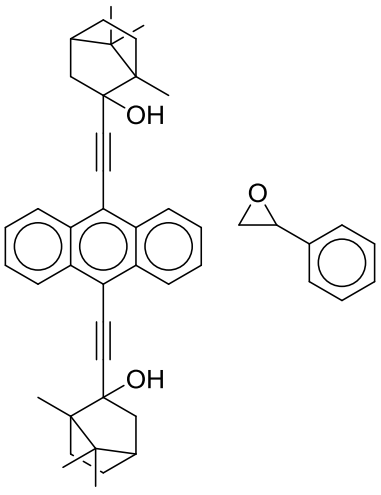
ZEPTUL



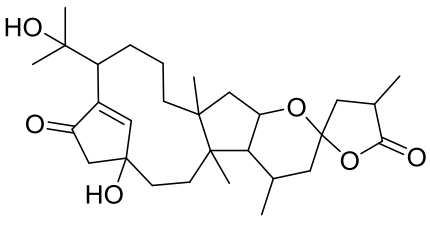
ZEQCOP



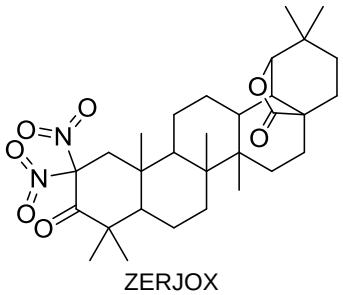
ZERCUW



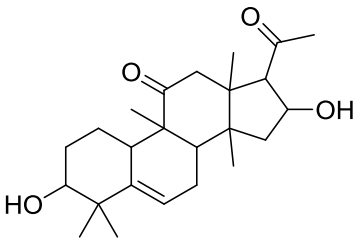
ZIHHAZ



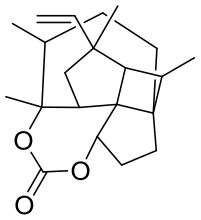
ZETNAP



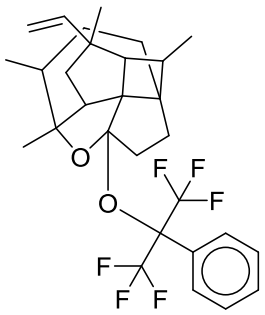
ZERJOX



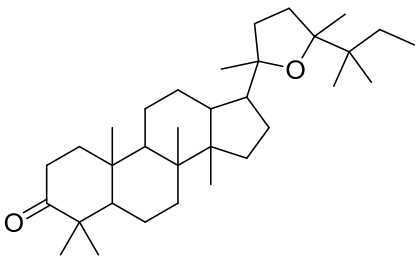
ZINRUJ



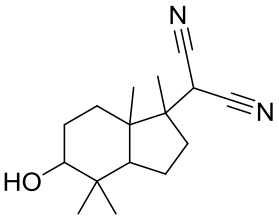
ZOPPUR



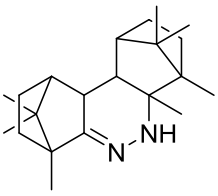
ZOPQAY



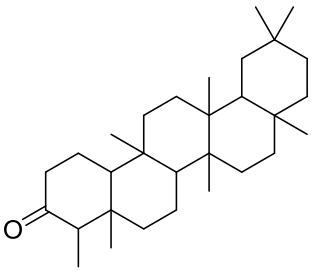
ZOXHUP



ZUFPUL

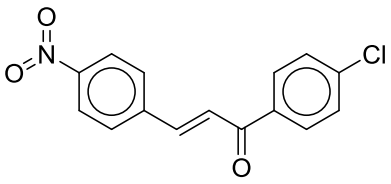


ZUMBEO

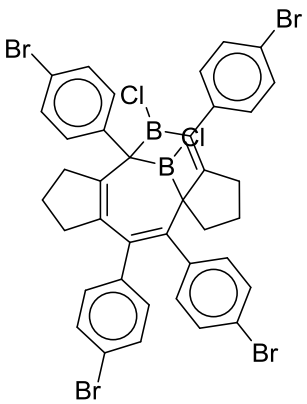


ZZZQAI02

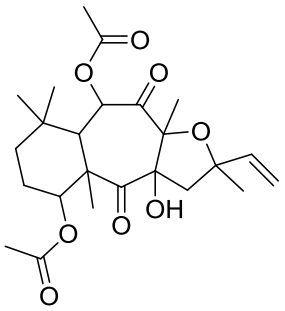
7.2. CSD structures with reference codes for Dataset 2 (1,002 structures)



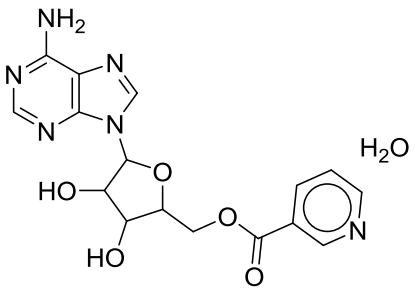
ZIJQEQ



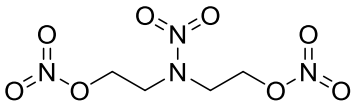
OQOFIL



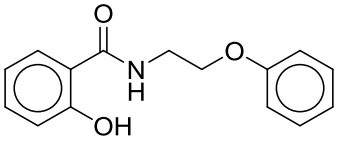
FEYCAN



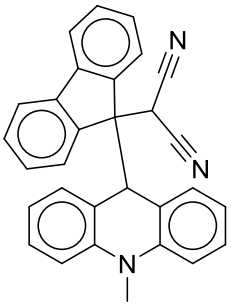
DUVHEH10



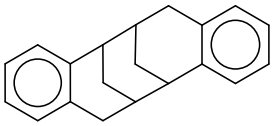
NXENAM



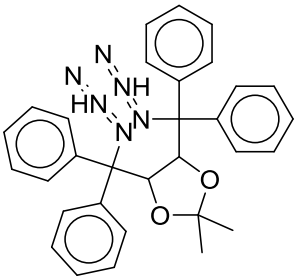
KEZKEG



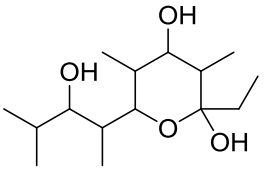
FAGQUA



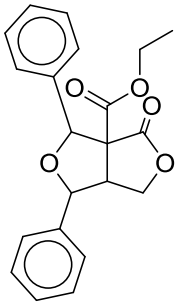
HDMBZD



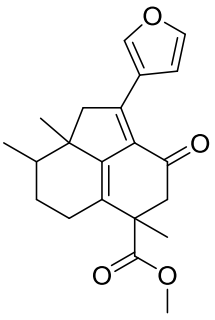
NIBZIH



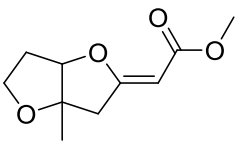
IDEVET



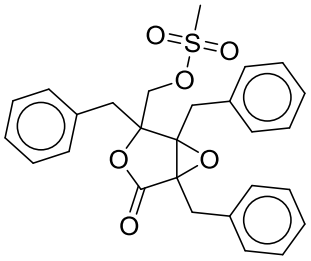
KALPUK



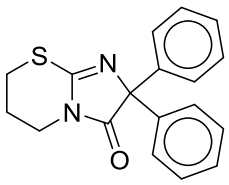
ODIRAU



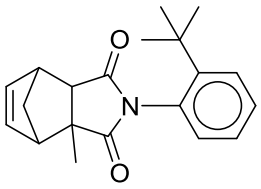
SUZVIT



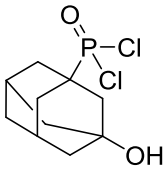
PIXNAN



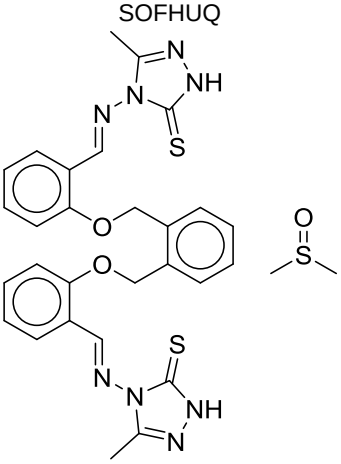
IMTHZN



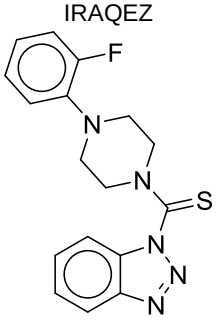
SOFHUQ



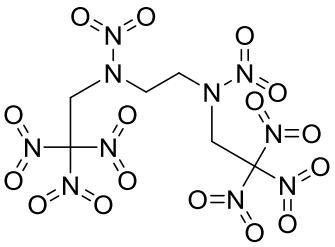
IRAQEZ



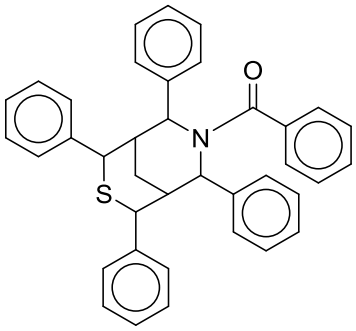
KIRPIM



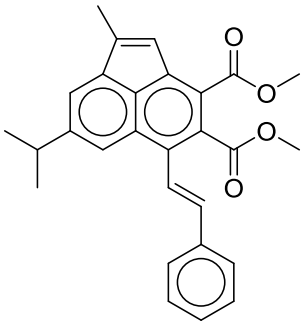
FIVXIT



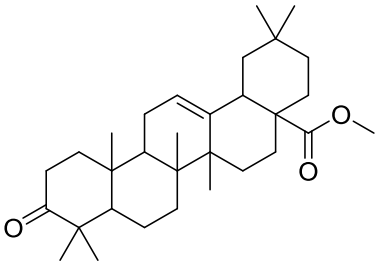
DILFUZ



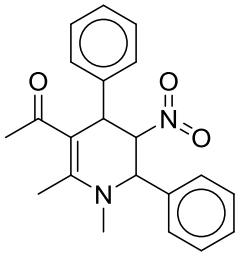
DUZLUG



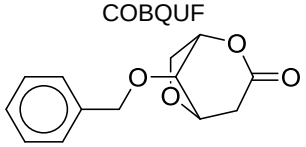
HEVZUD



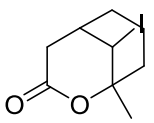
COBQUF



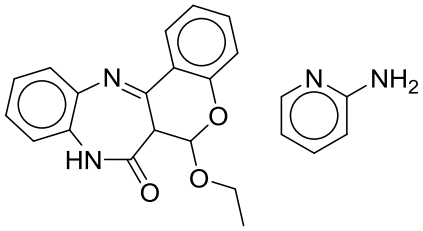
COZYIB



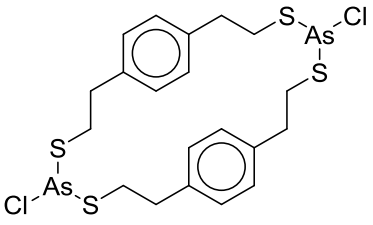
QIBLUI



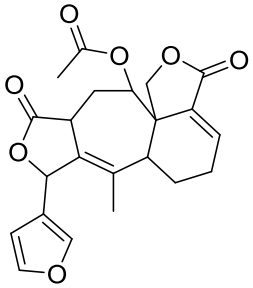
QIKHEZ



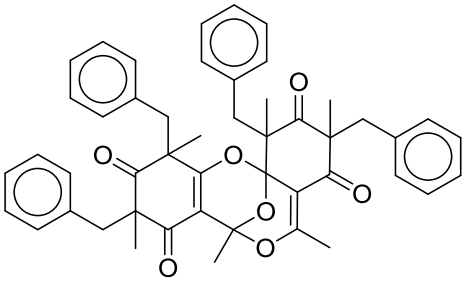
ZUFQAU



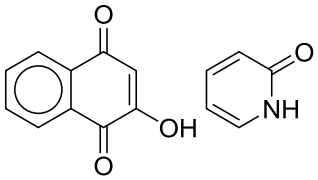
FUWXEB



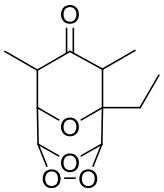
DIDHIH



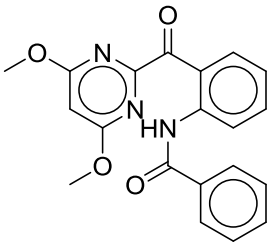
ZEJJIH



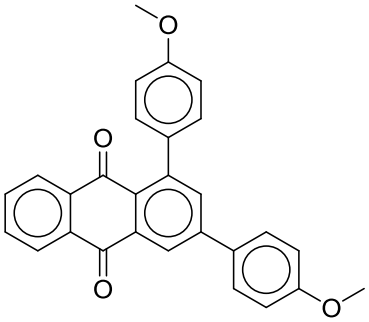
GUNWUJ



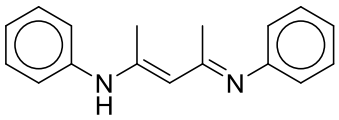
TOZQIK



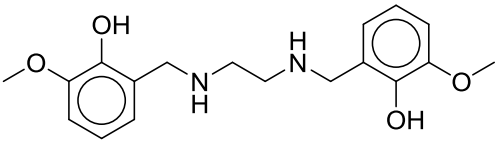
HIRMAX



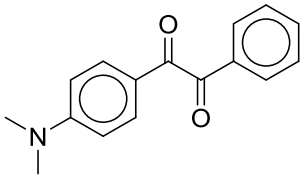
YIPPAQ



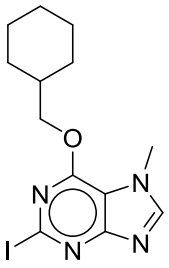
CADSEF



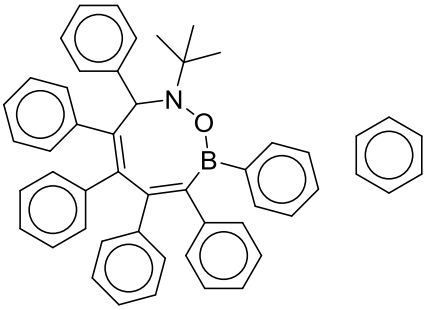
METYUG



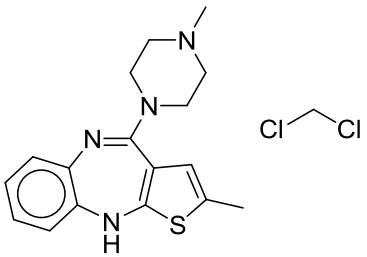
VOLLEO



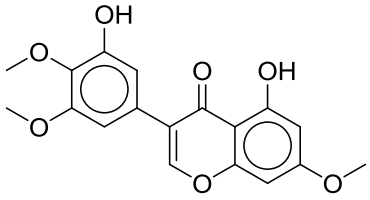
PEXHOR



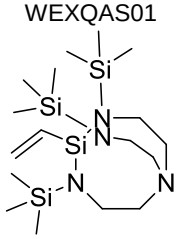
EQEHEP



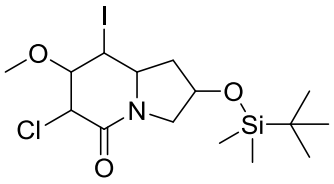
WEXQAS01



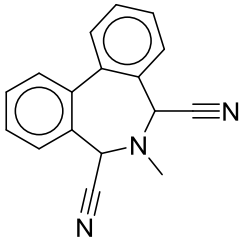
SUZRAG



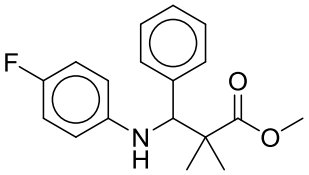
AQIWAY



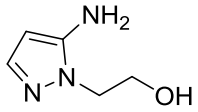
ROGNAD



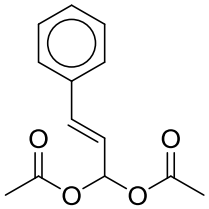
WIPJEK



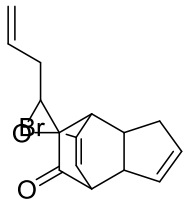
PODGAR



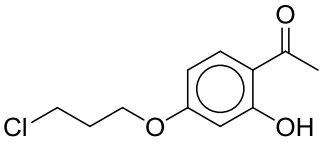
KEQXIO



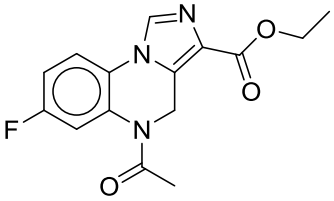
NABLEI



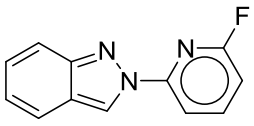
MUBBAM



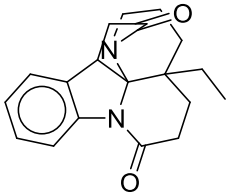
HUFDIW



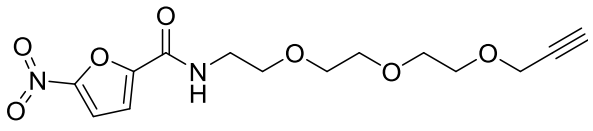
EKUWAK



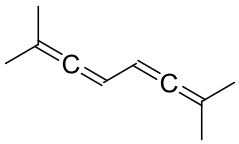
DUDCUC02



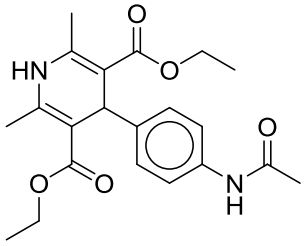
XIDMIH01



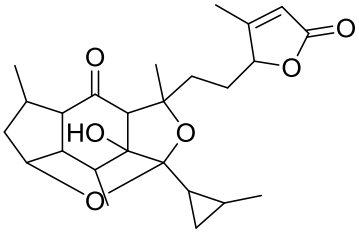
HOHVIL



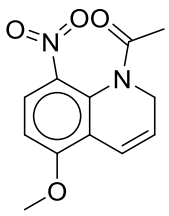
FACXOX



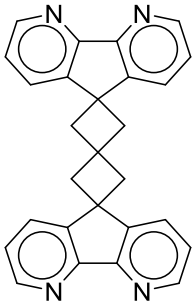
LEGJOY



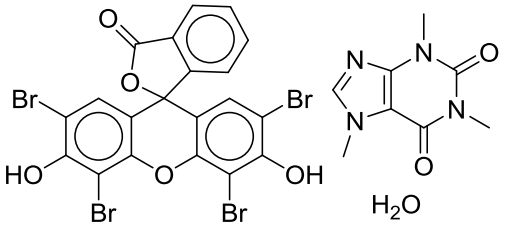
KEWKII



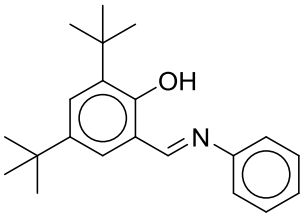
HUTKIR



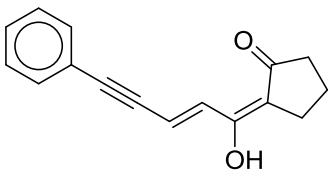
NAMMOE



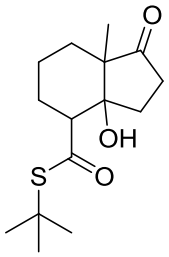
AGIKUY



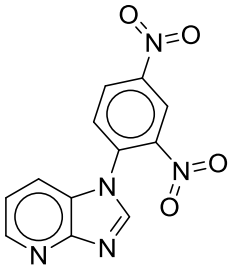
IVOHIL



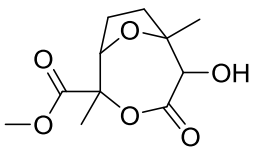
VUQTIM



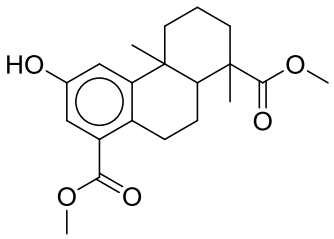
SAYNIR



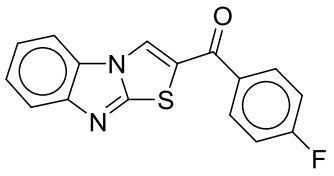
NPIMPY



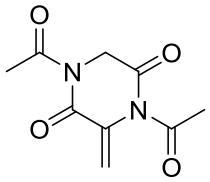
YIZNOK



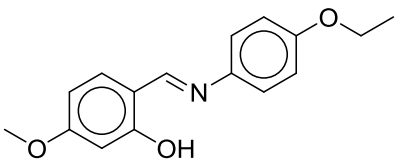
ECADEU



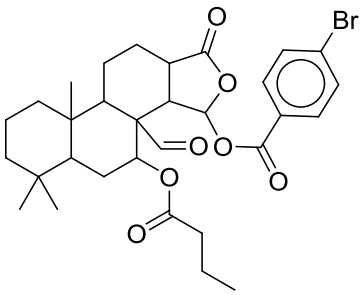
ABESOS



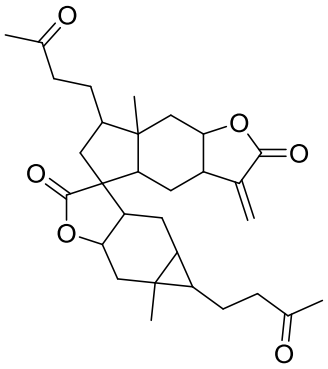
HEJFEJ



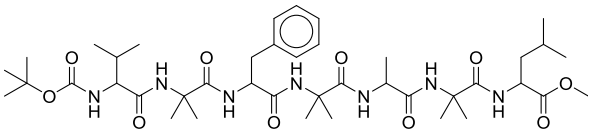
PUGHUV



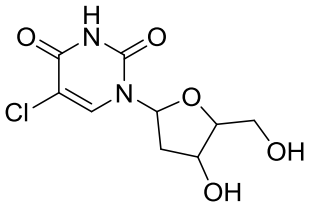
NAYMIJ



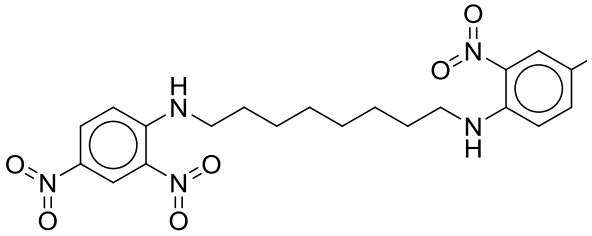
GUDQED



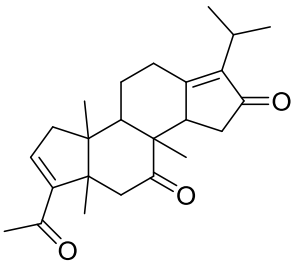
LUYPAW



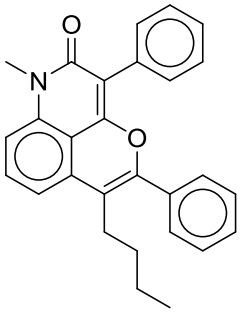
CLDOUR



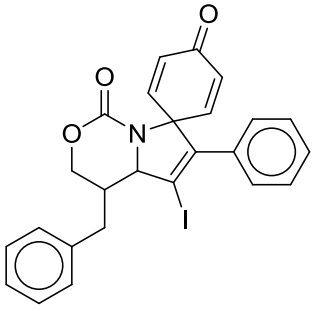
QOVTEB



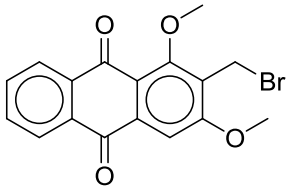
OTIJOS



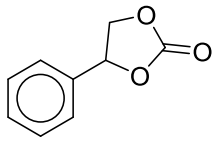
XIGGAX



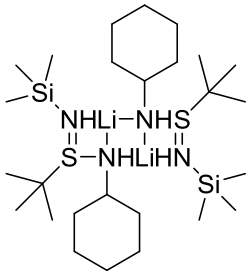
UMEGAW



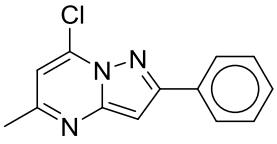
QOHSIR



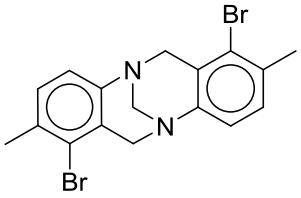
ITELOK



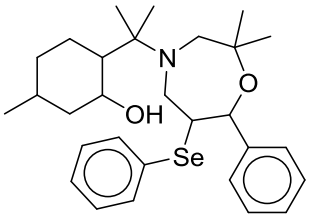
WECWUW



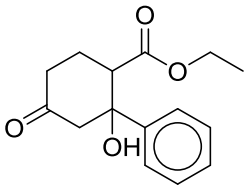
FILZOR



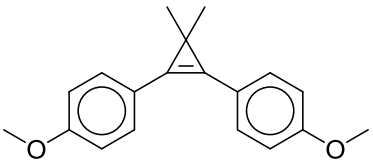
JEVYOZ



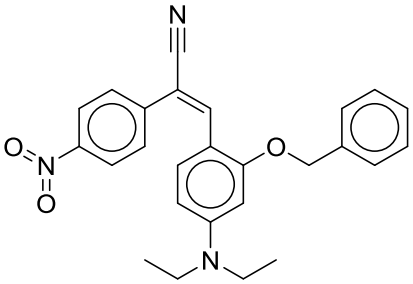
BULWUC



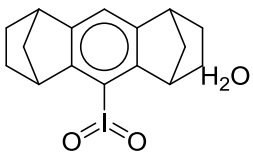
QIJYOX



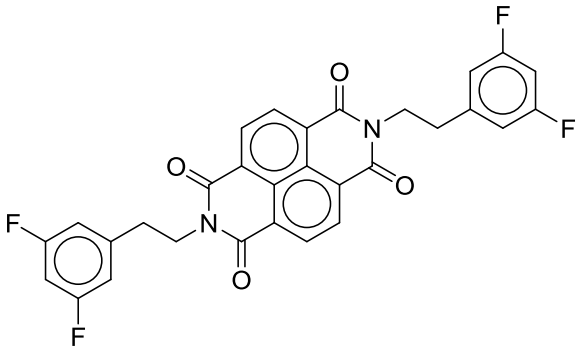
YONTAX



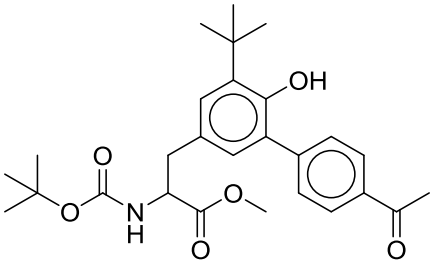
ANUDEU



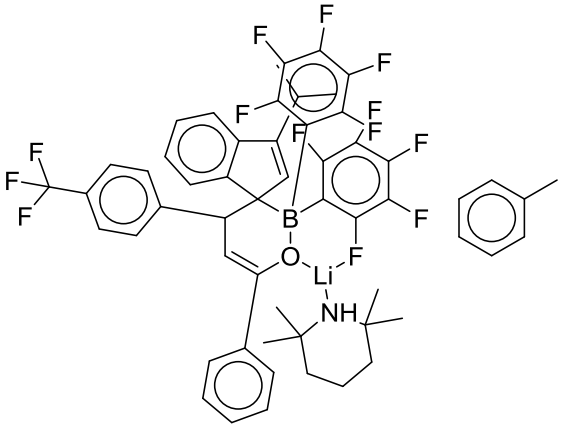
FAYWOT



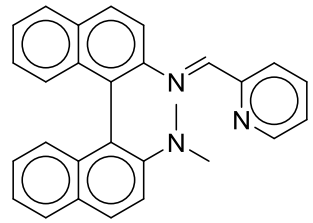
BOSKEA



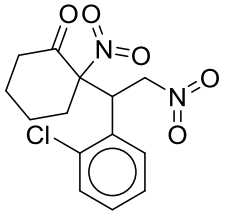
GUGZIS



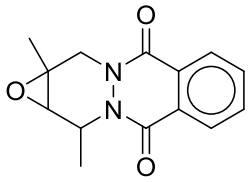
DUYYED



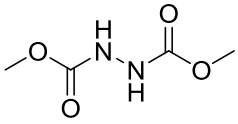
ULOGEI



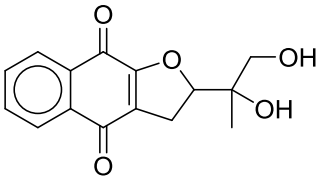
BEGREM



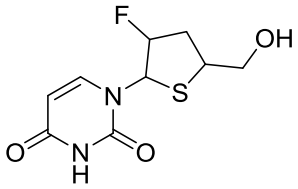
ETZAND



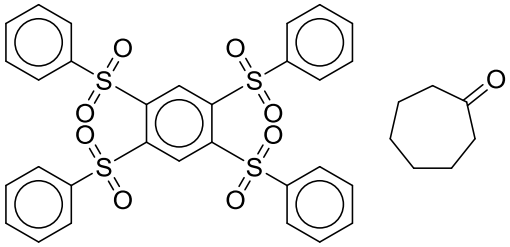
JIPFAQ



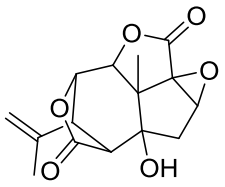
WUQKIE



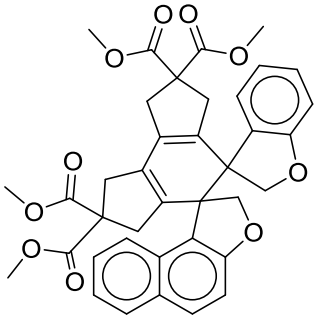
YERDAA



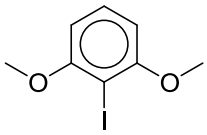
ZEKQEN



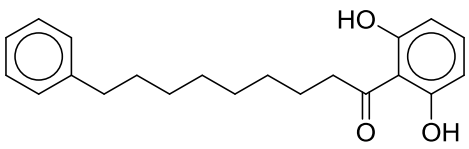
CIBCUL10



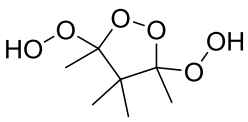
VUXBEX



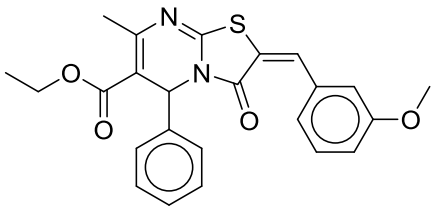
MUBQUW



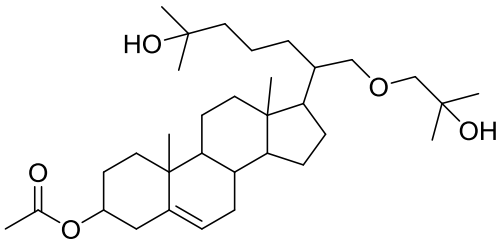
WEFMEA



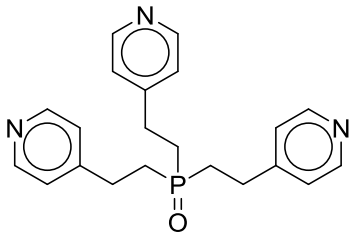
YAFNOL



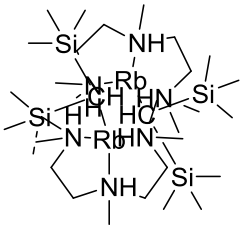
PUJRIW



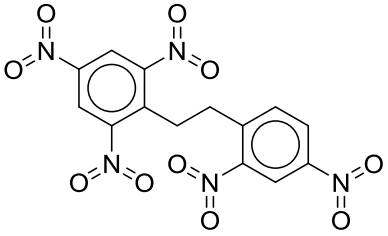
LIQBUI



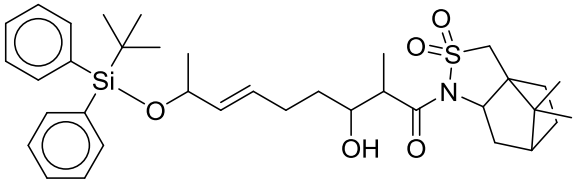
POTRUN



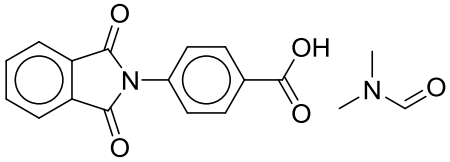
PETYAO



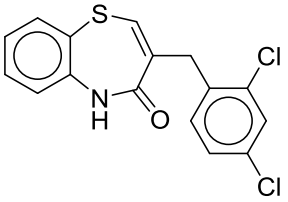
YAGHUL



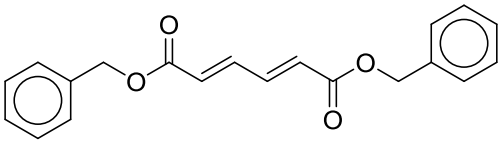
BACVUX



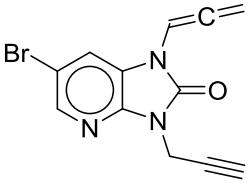
YERTEV



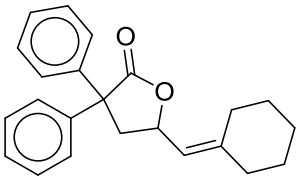
WIGRUB



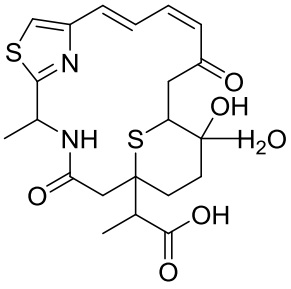
MONTIS03



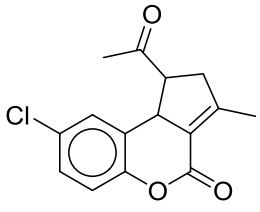
KUSPIY



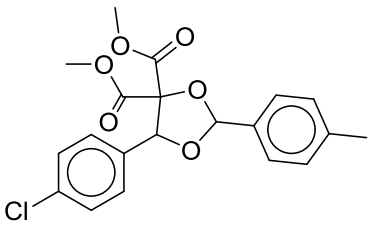
COTDIA



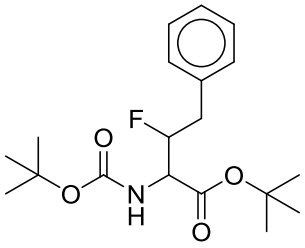
QUJYOL



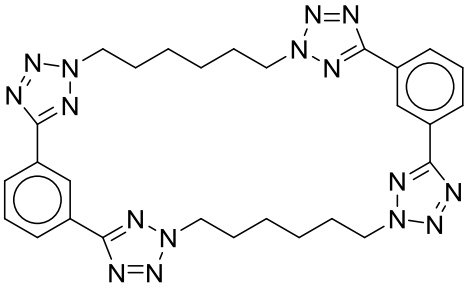
PINCIA



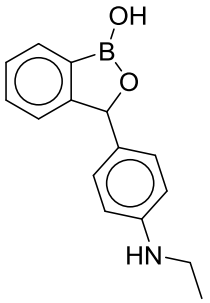
GEBKEF



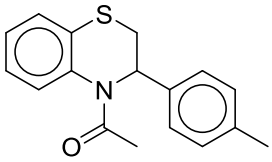
FAWJUK



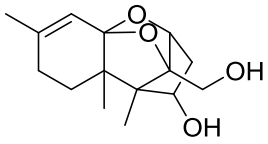
WIFNIJ



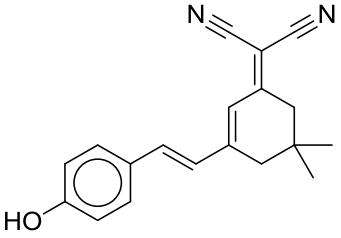
WUXHON



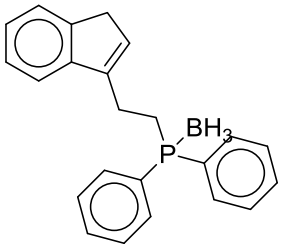
FUNCAU



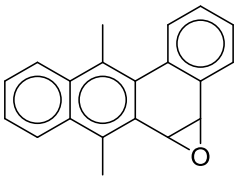
BAMJEH



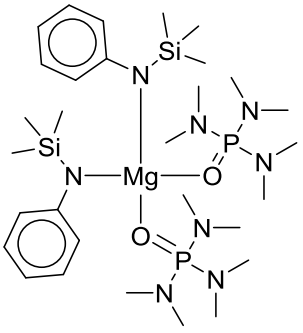
ICAPOR01



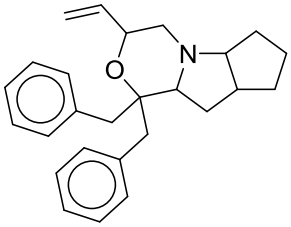
WOCBEV



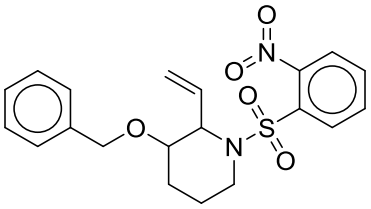
MBANOX10



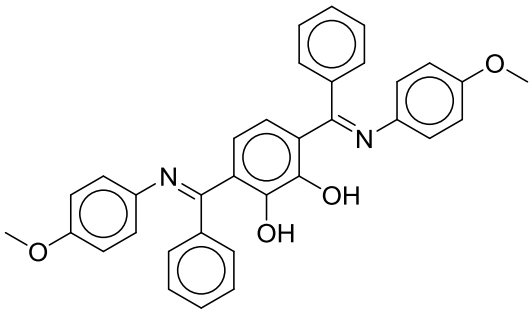
AQIBIM



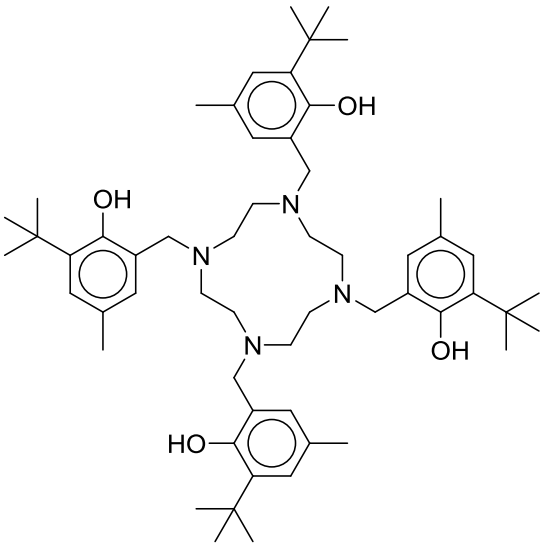
ZEWQUN



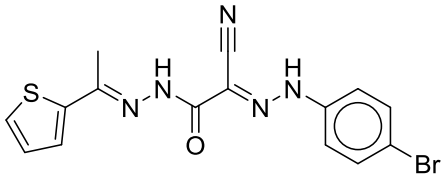
UBAZOO



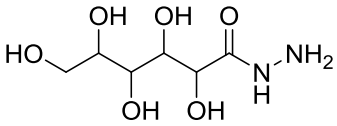
COCFAC



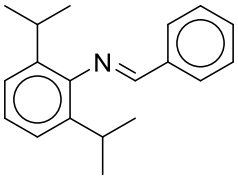
ASISIG



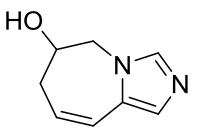
XAFFIW



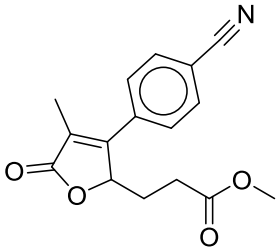
LACBOG



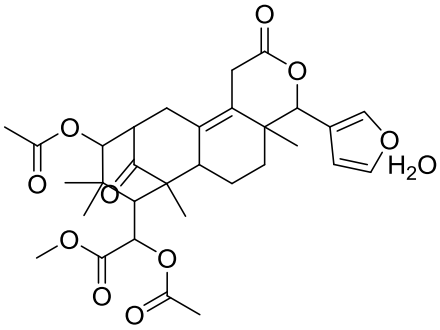
GIRJIC



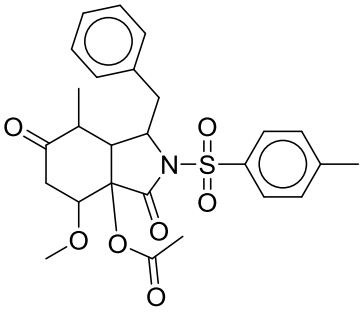
VEYPAR



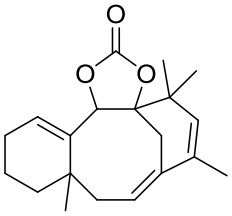
CUSJEG



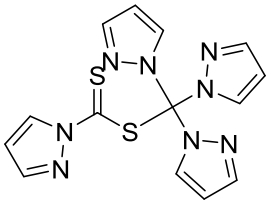
HEWTOT



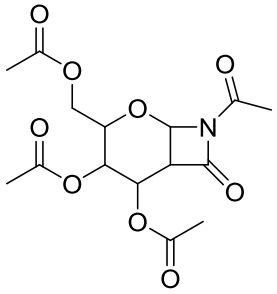
GEZLIG



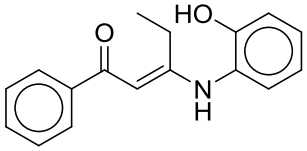
OKUCOO



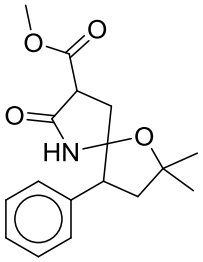
COKYEI



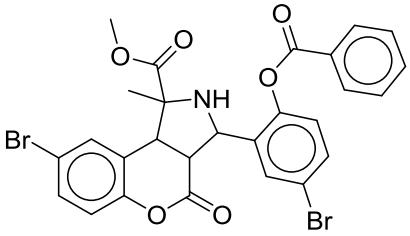
TEWLEM



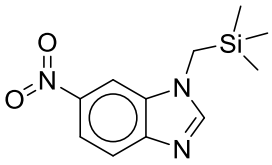
KIJPEZ



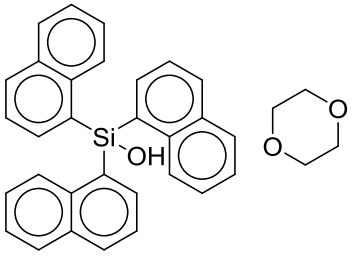
MIZDAB



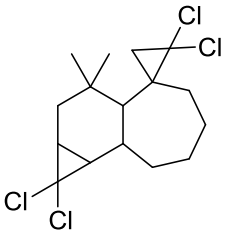
IZIREQ



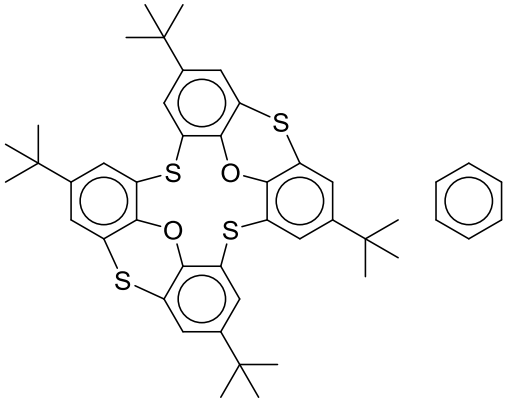
ANOXEH



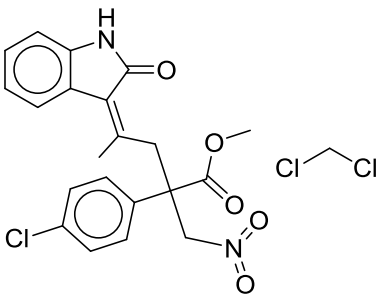
JODYEG



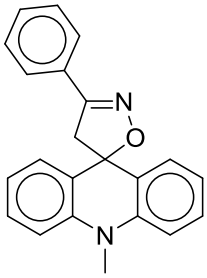
SILYIV



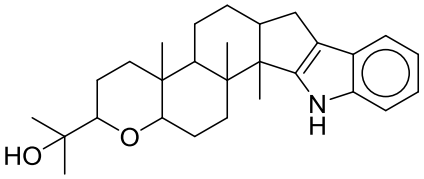
MOHTIN



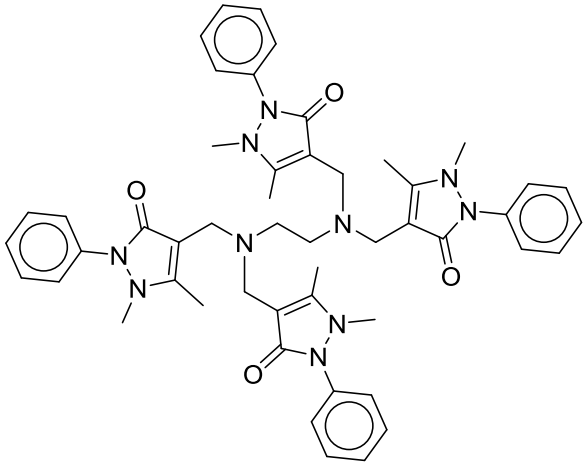
RUNQIC



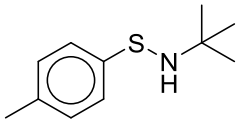
IYIKUY



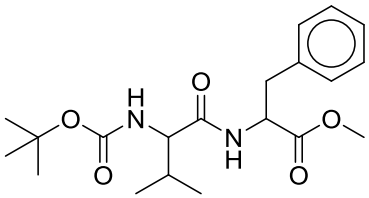
HUHYUG



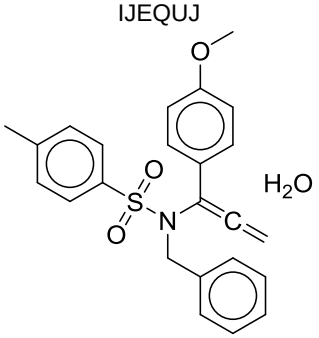
NAXXAM



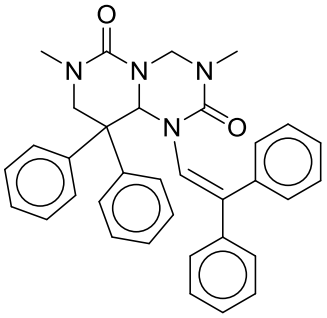
IJEQUJ



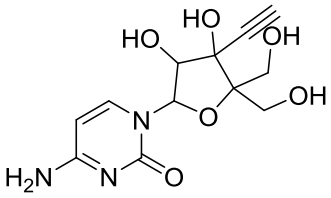
PATWEN01



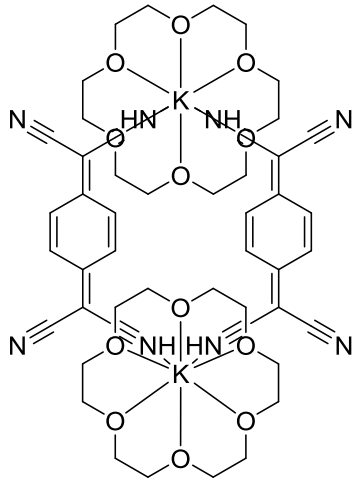
PELCOA



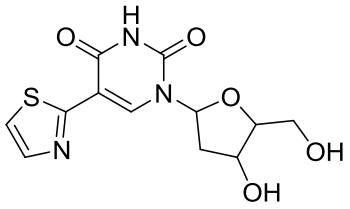
WAHFOB



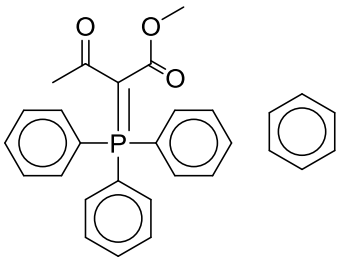
FIYTEN



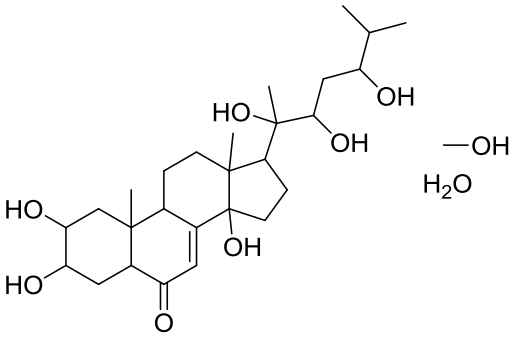
JEXJAX



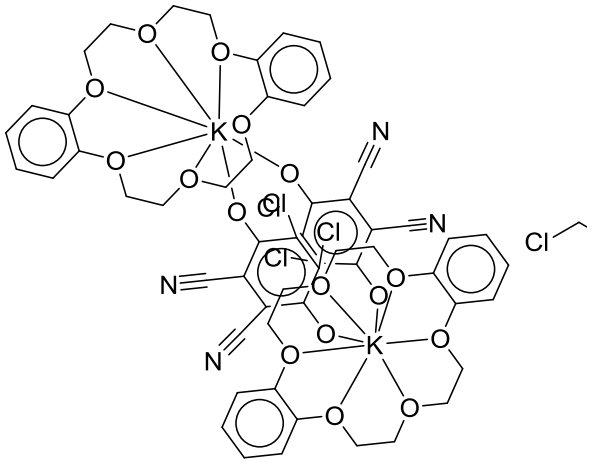
HEZWOZ



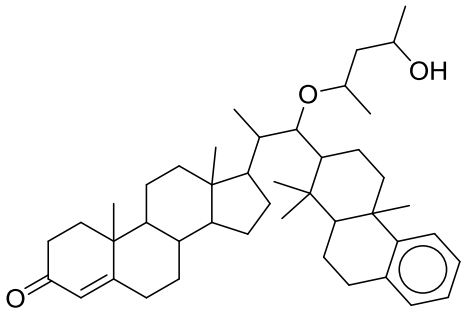
GUKTUB



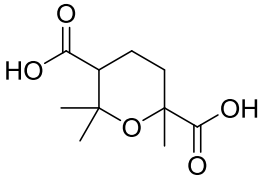
SONQAN



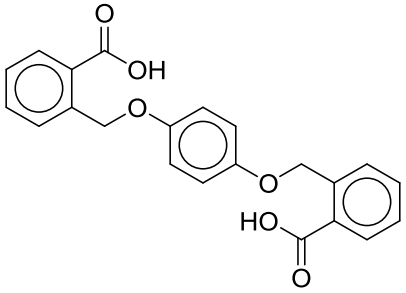
DOVWAN



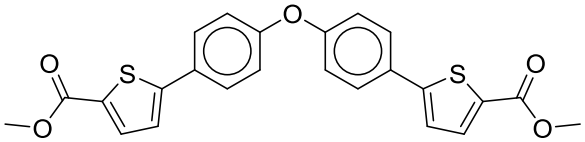
VIXNIA

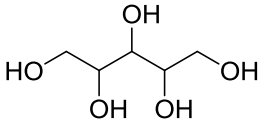


TADNUI

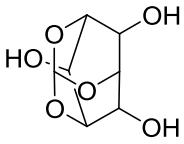


UHELOI

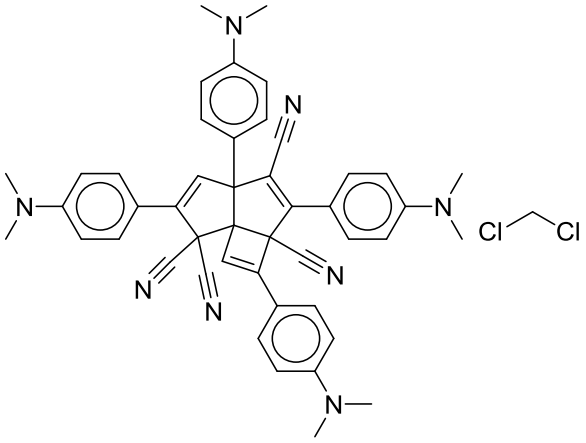




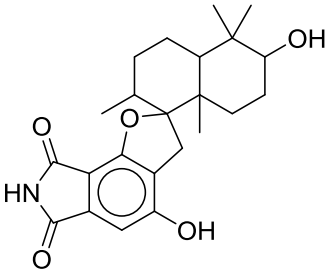
BEQYOL



ARABOL

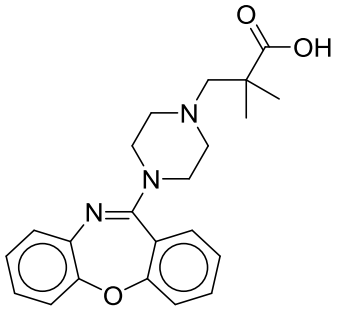


PARGIY

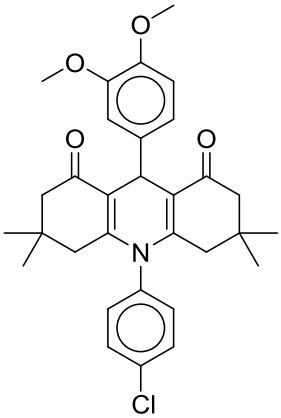


H_2O

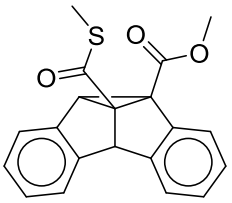
TIWFIQ



LIVWEU

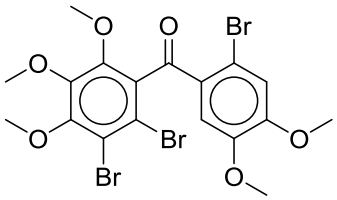


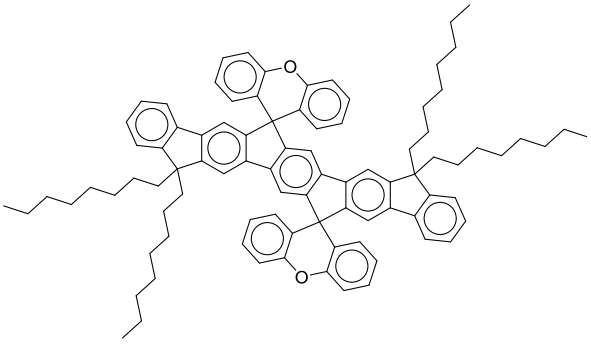
FOCWUR



CACYIR

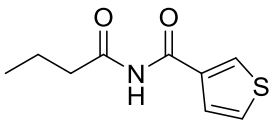
ZILBAX



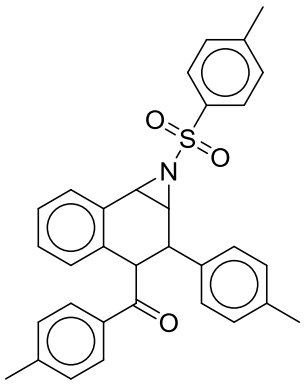


ASEXOM

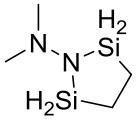
IBUPEC



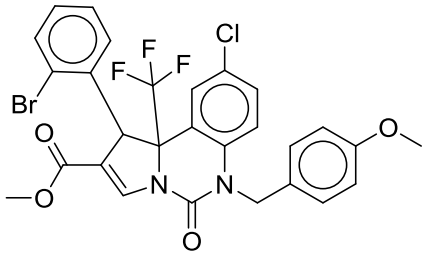
UMILEH



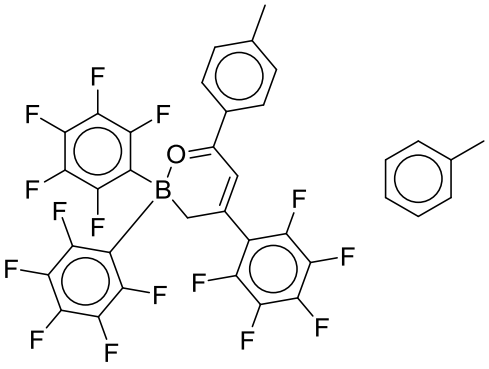
JUBBEP



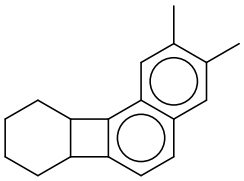
RODVIP



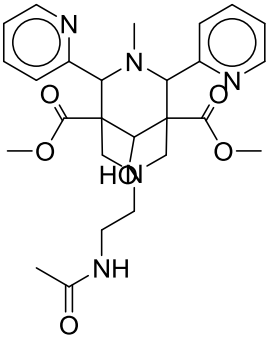
YIGYOE



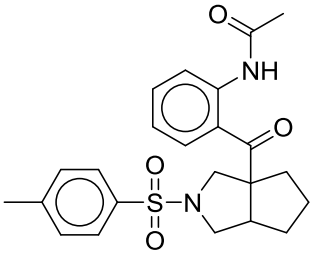
OZUYIT



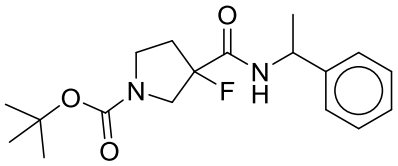
BEFJON



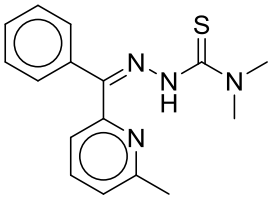
WUJVII



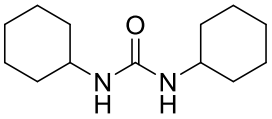
UCOQUB



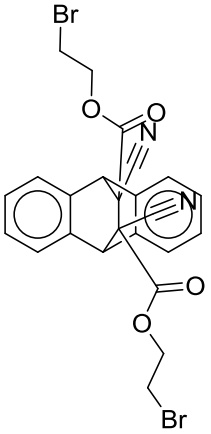
OXUXIQ



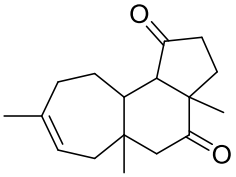
XETPET



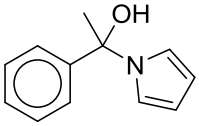
CYHXUR08



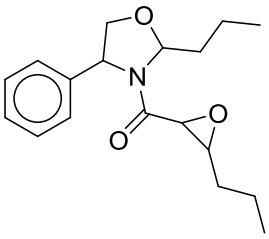
VOBTIR



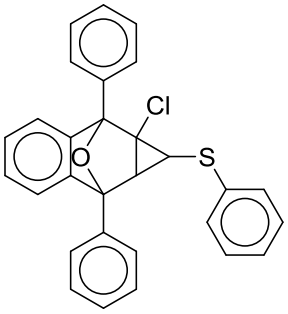
IFOSEC



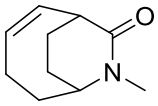
EHAQEJ



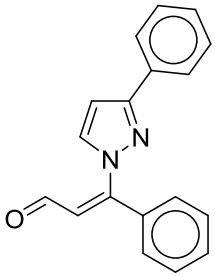
PIHBEP



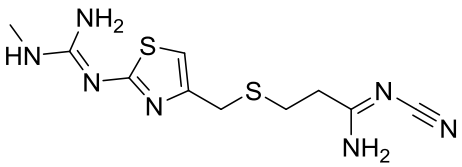
PODJUO



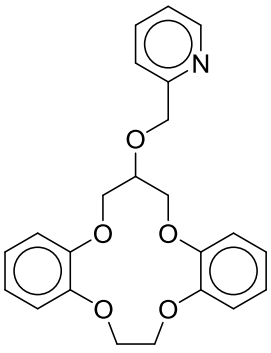
SOJXIZ



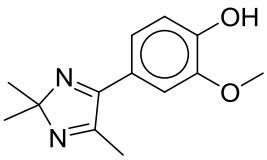
PUJNEO



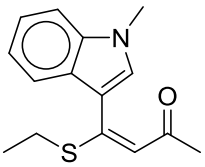
JATMAS



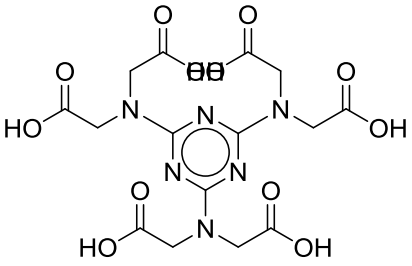
SOGQUA



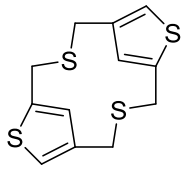
MAZDAU



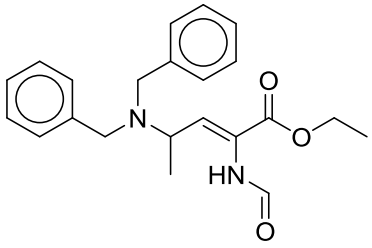
HOYPER



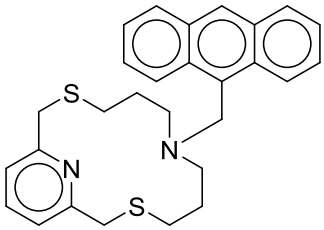
LOMMEF



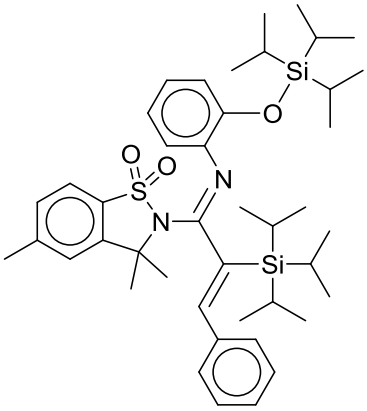
LAFFON



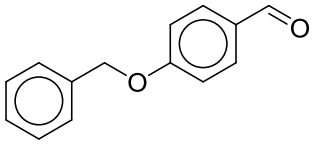
PADXEX



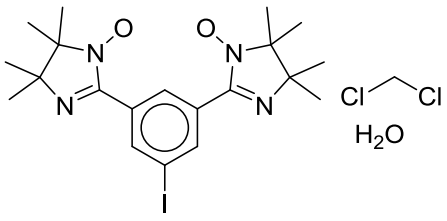
QAWJEE



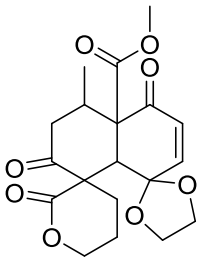
FURZAU



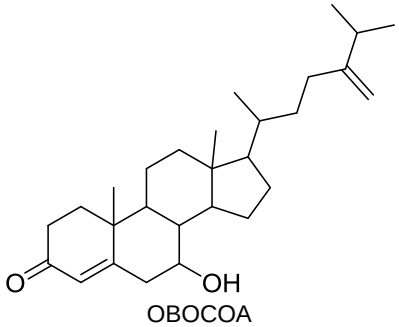
DUTRIU02



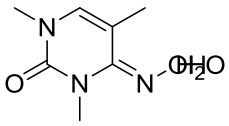
QIWDIL



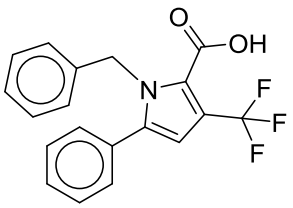
BANFIH



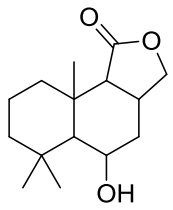
OBOCOA



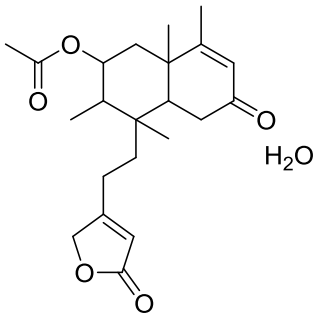
NETGEY



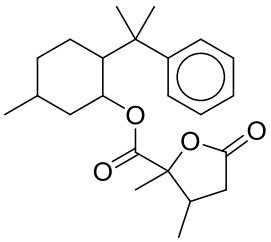
PUBXIT



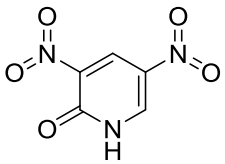
BUHKEV



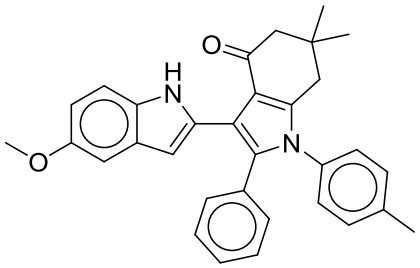
VACZII



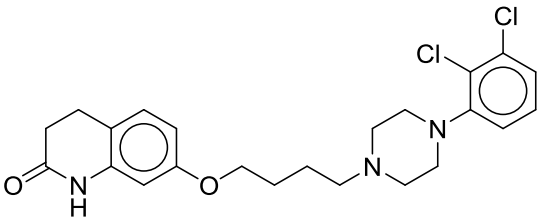
PACSER



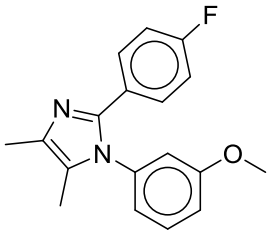
JERVUY



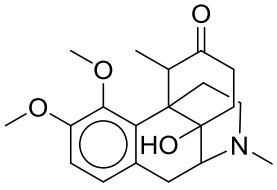
NOMQEN



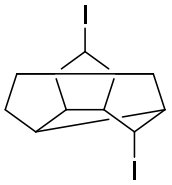
MELFIT06



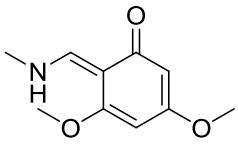
RIPQIS



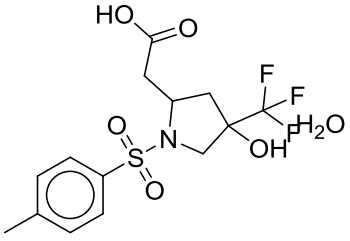
AKILOV



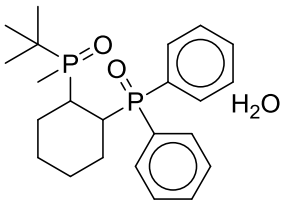
IPCUND



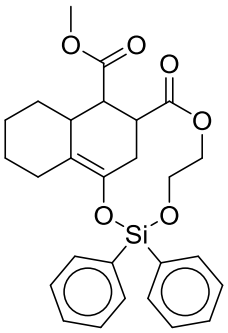
QEVSEQ



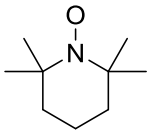
WOCNIN



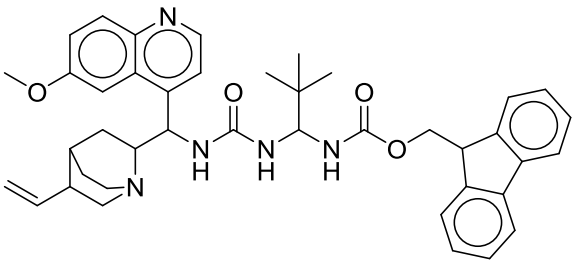
WAFJUL



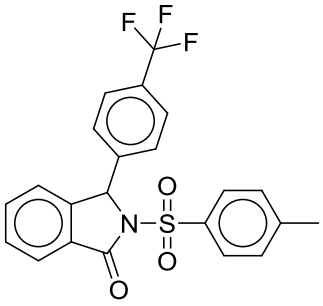
SOJHUU



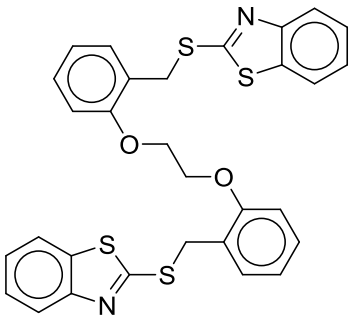
TMPPIO13



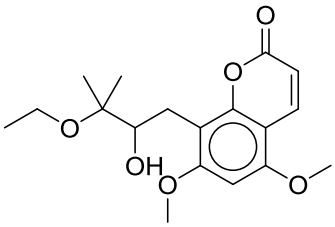
VIJDID



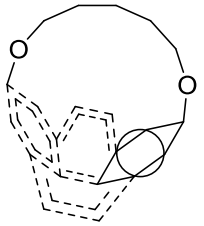
REMHIC



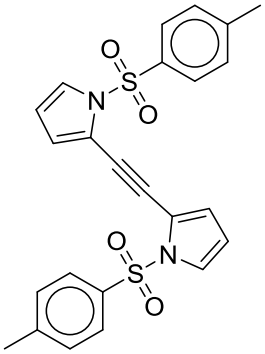
SUYTUC



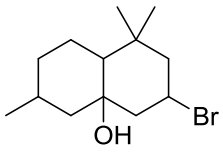
RUCYAR



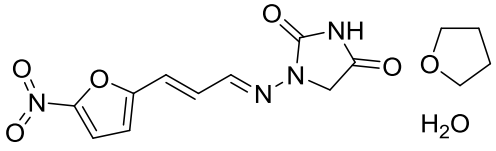
LIVTAL



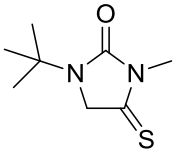
GIRRAB



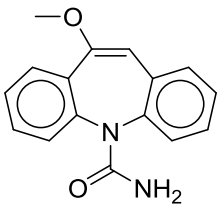
WEJQIM



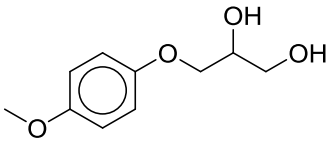
ASATIZ



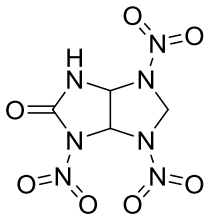
GUQHOR



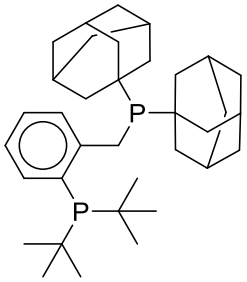
FONCAN



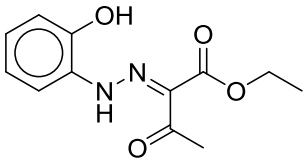
DAMPAL



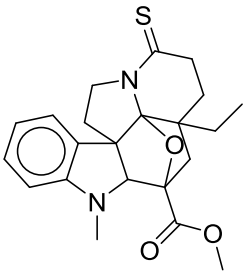
ABEJOH



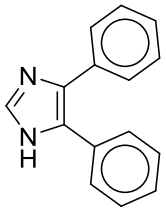
CINWED



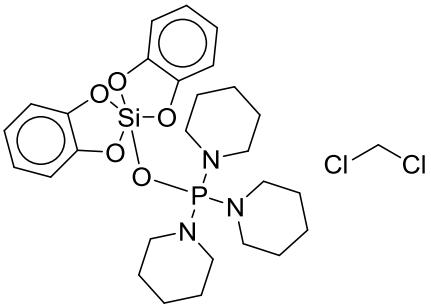
ITIJOL01



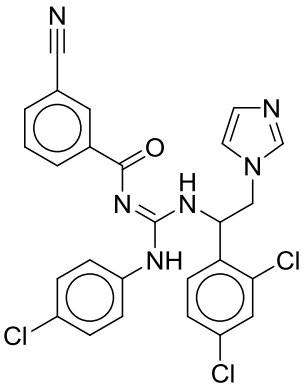
VEQDOL



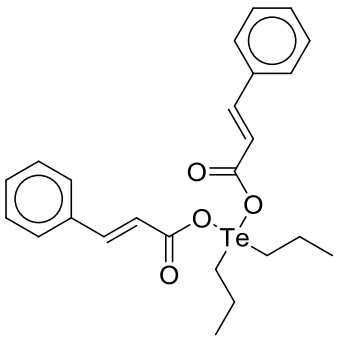
OCUSUA04



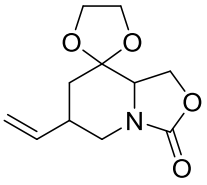
VUSBIU

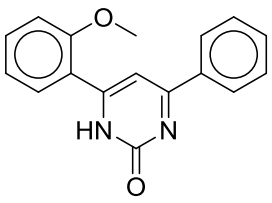


SAQKAX

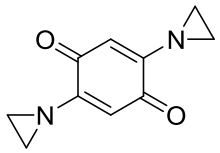


KAZDIA

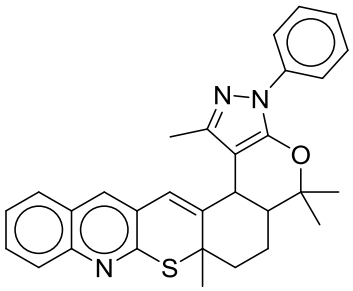




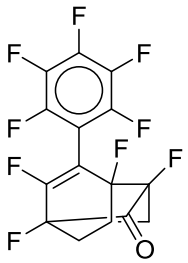
LENTEE



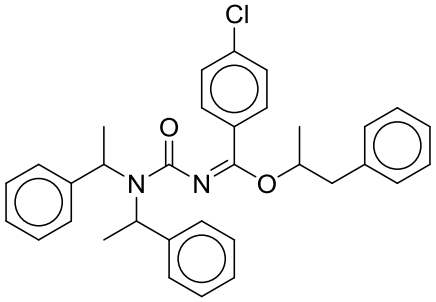
UCEWIL



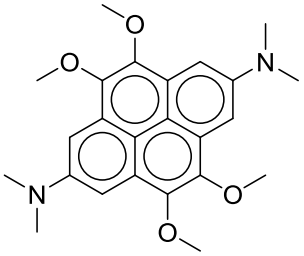
ETIMQO10



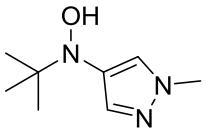
GIHTAU



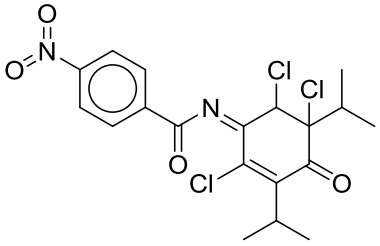
BAQYAT



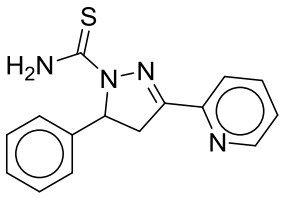
LAPWIJ



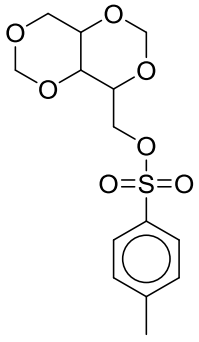
TEWXAW



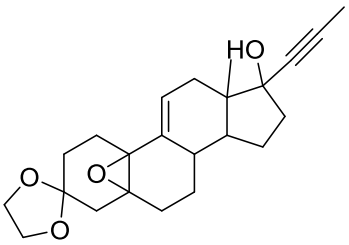
HECLAE



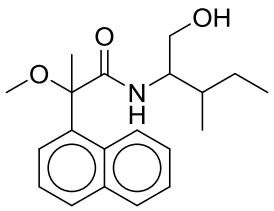
SOHFEB



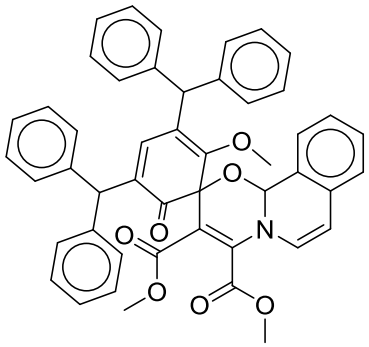
SERPAI



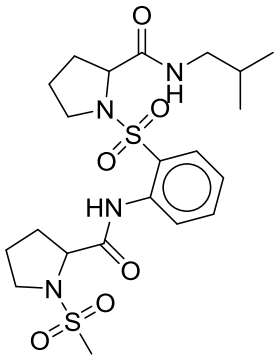
HALSAO



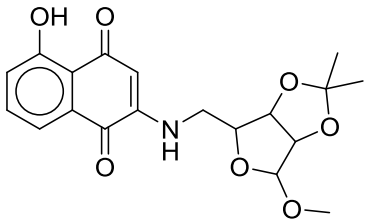
ZENNIP



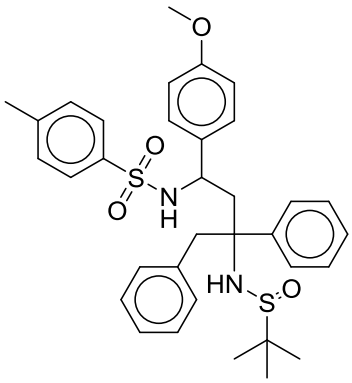
YIHKIK



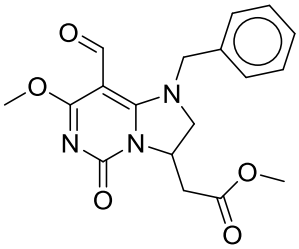
LUVSIE



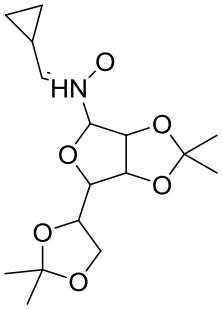
MOXHIS



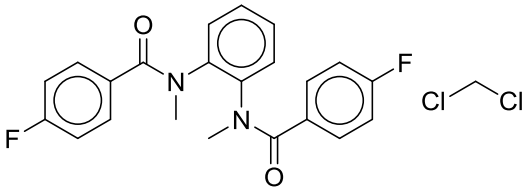
YUYHAD



VEJQEH

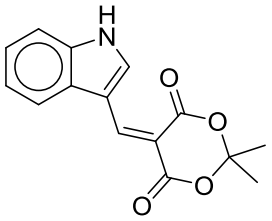


GECQOU

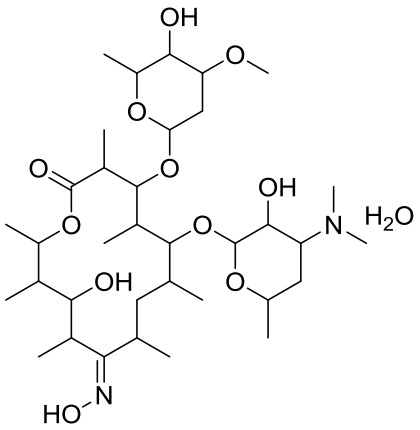


OGUGOM

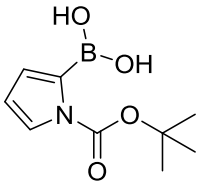
TOHVUI



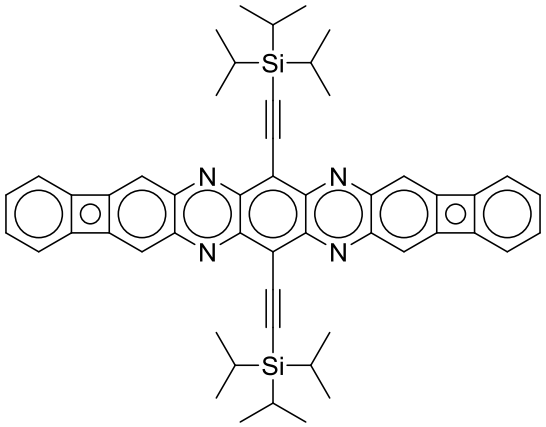
EVIBIV



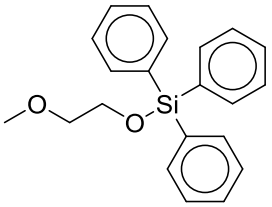
LIDBIJ



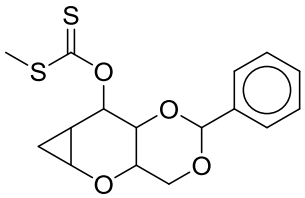
NIQCUM02



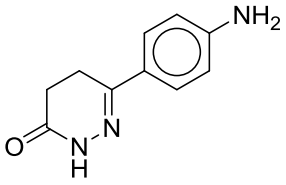
UZENAQ01



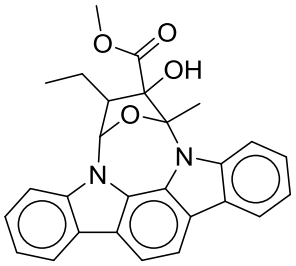
MUXLEW



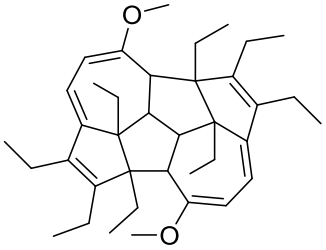
XUHPAR



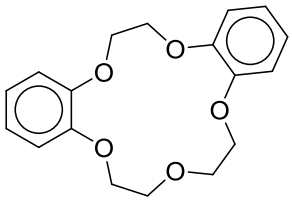
HAVJUJ



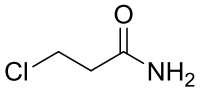
QUFYAR



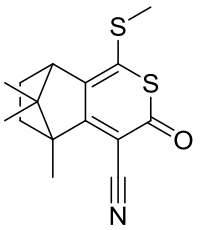
JEDAEY



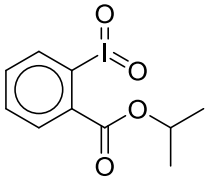
SOJCID



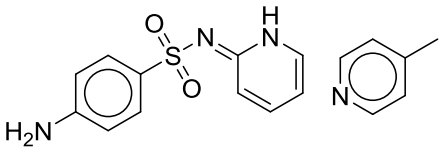
WOCJAZ



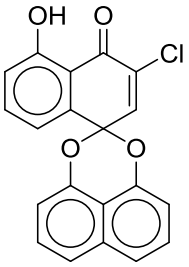
POTGIO



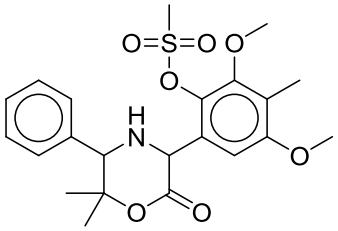
INUZOG



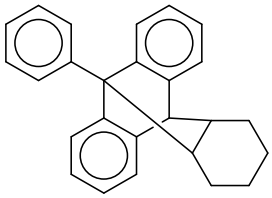
XIFLAA



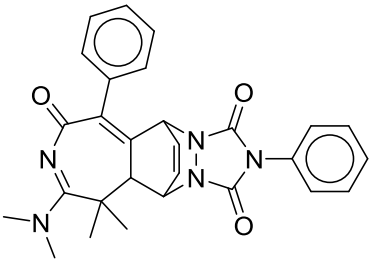
YOGWOI



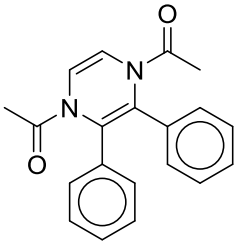
VODNOR



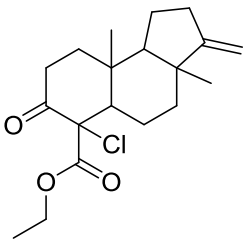
GUHFEU



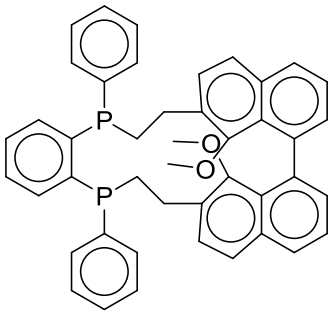
MAEPYZ



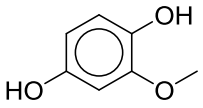
BIDDUN



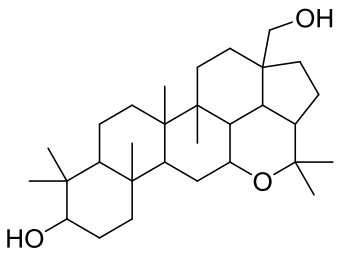
AHUDOY



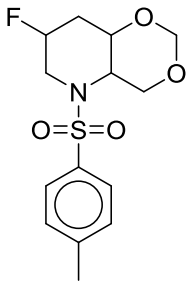
JOPCIA



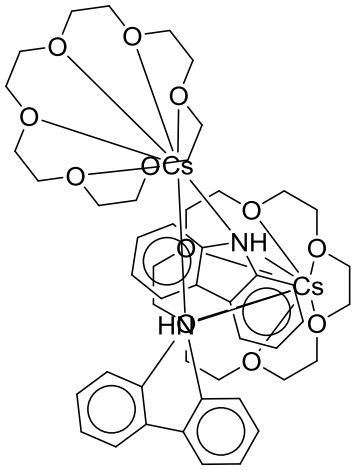
DIFDUS



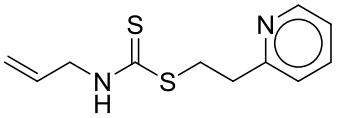
KOWBUV



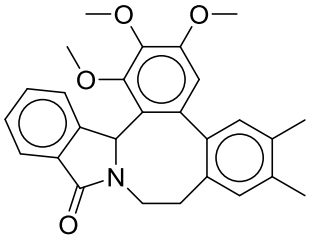
APIPAS



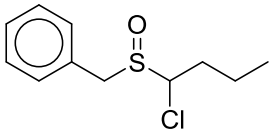
XAQSUE



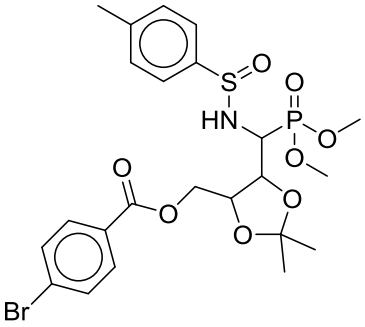
AMAYUK



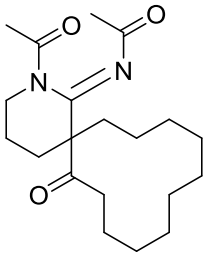
HESZIP



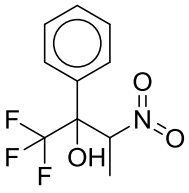
PIWCON



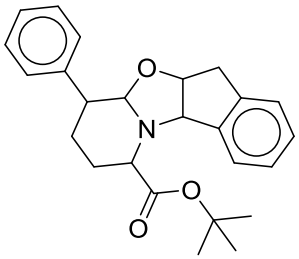
OPIBEV



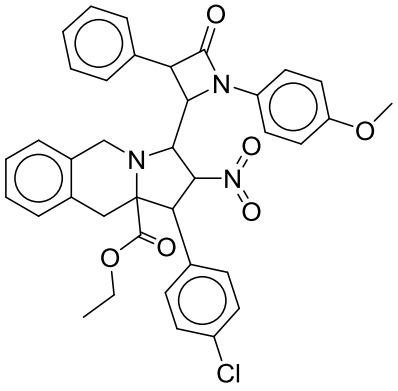
ZOGRIW



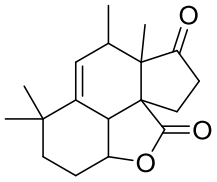
VAGVIK



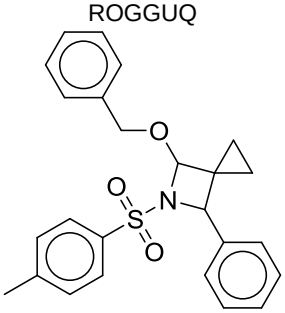
ITAZUA



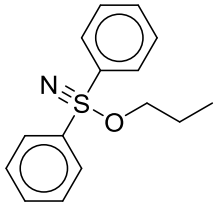
ROGGUQ



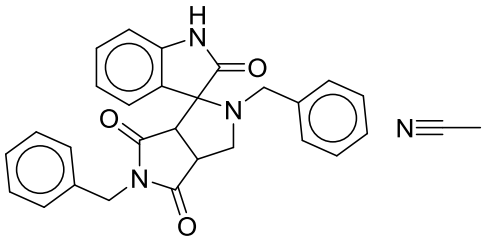
OCETIC



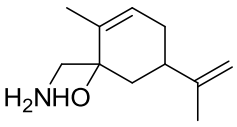
LEMCIQ



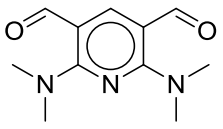
NUYGOD



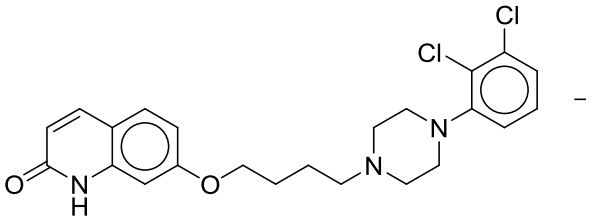
IRINAY



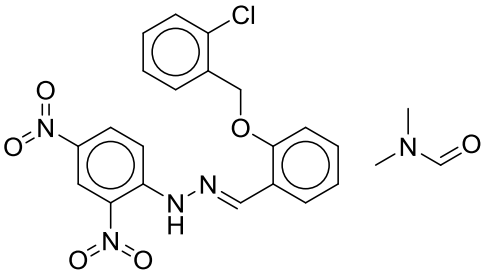
WUCZAX



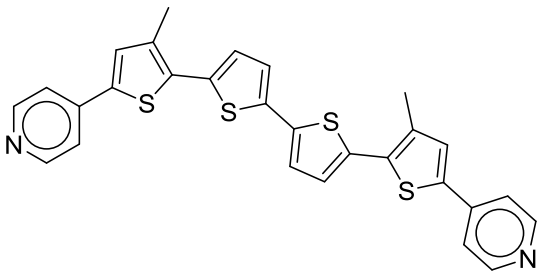
WIKBEX



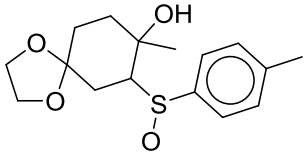
QIFJUM



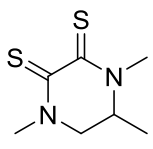
NEQLAX



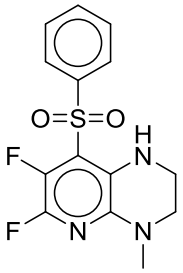
CADGOF



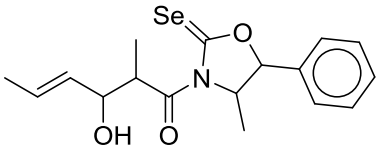
ZAPWIW



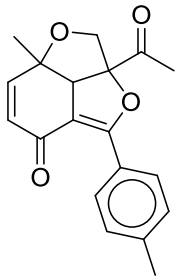
KIZLEL



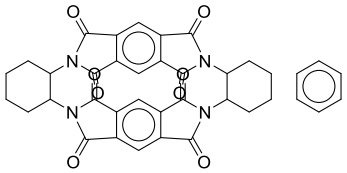
MAYVUE



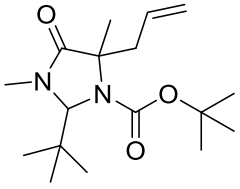
WONQOF



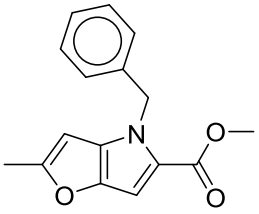
UDAZOQ



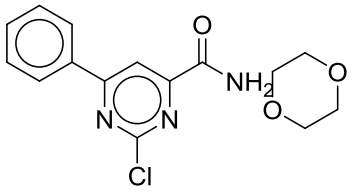
MOLMUV



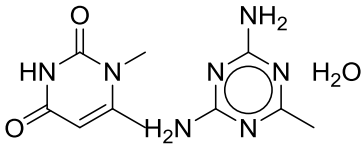
NAHSIZ



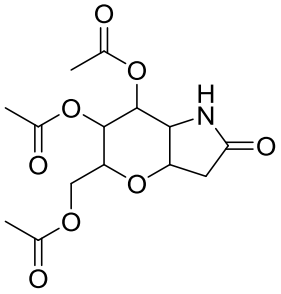
OCIFOV



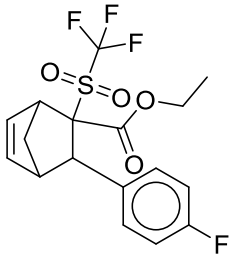
TITJAI



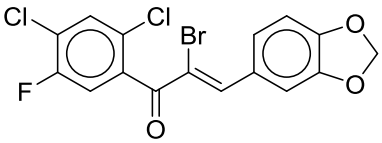
HIMMEW



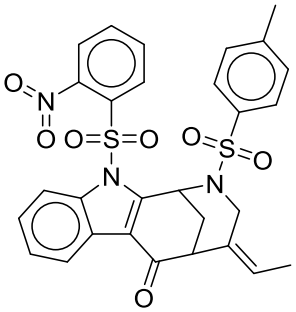
EBOKUE



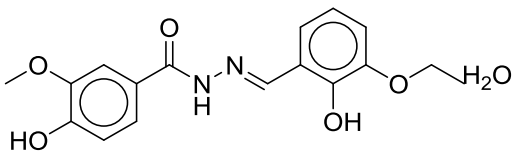
DAQTIY



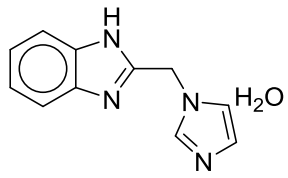
RIBZUY



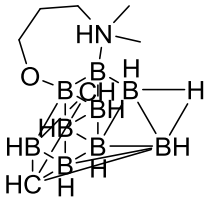
ILAJEM



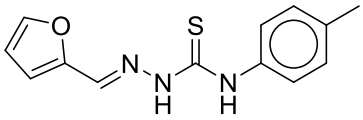
BUDKOB



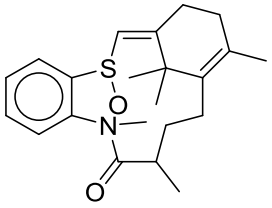
LORTUI01



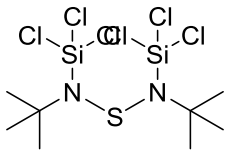
MOCPEB



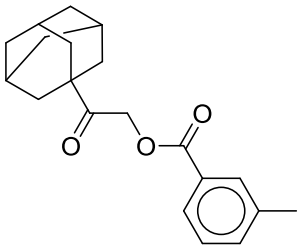
IJUFAV



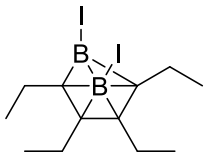
MEHLOA



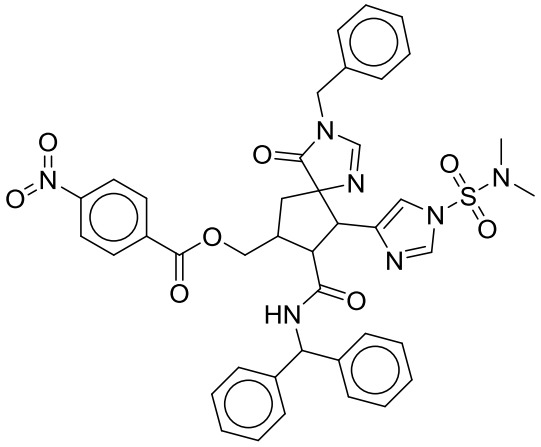
JIZWUK



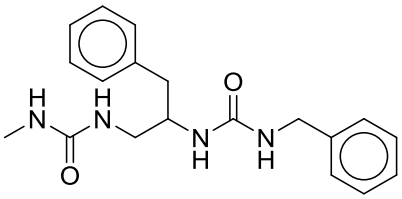
BUVDON



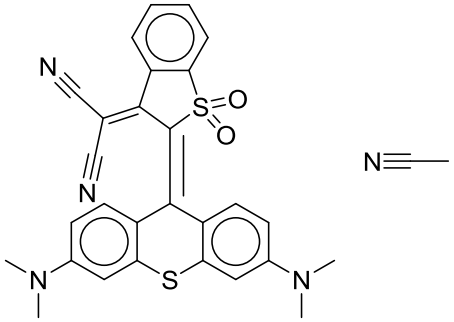
AXEPIC



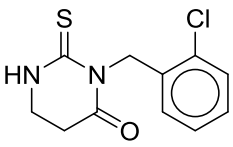
LIPSIN



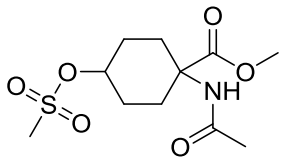
ROCPIJ



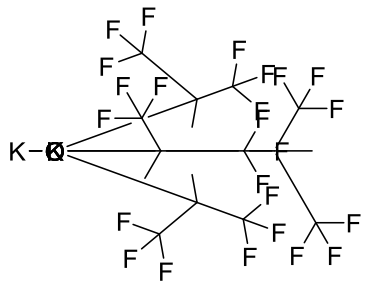
WUGLEP



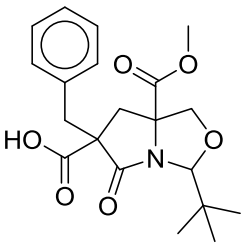
EWOTIT



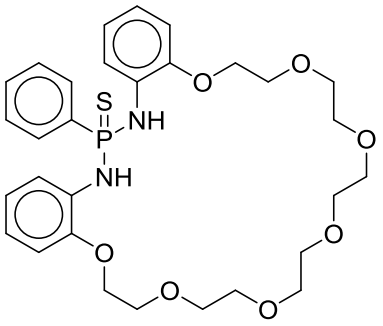
DECQEH



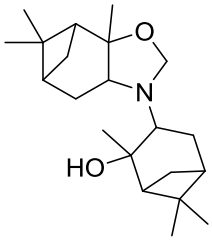
UDIJIC



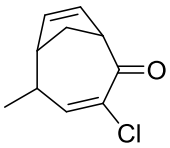
ECUJUH



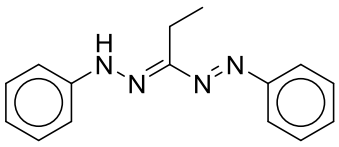
EDOYOM



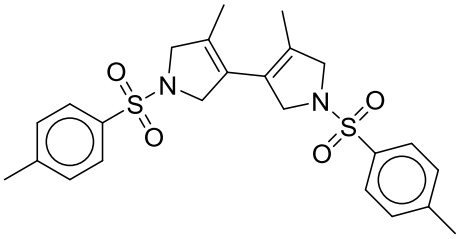
RIRLUA



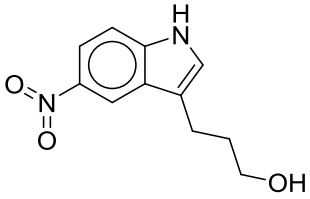
HACJOK



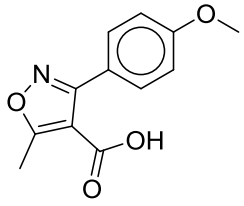
GELZUS



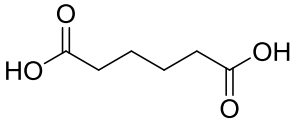
WUNGES



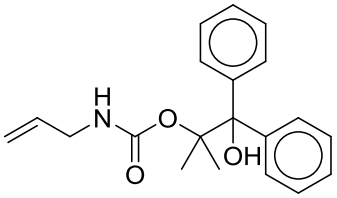
SUVQOP



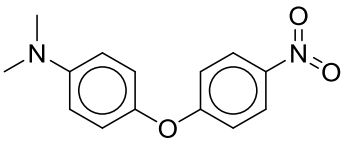
MEWRIR



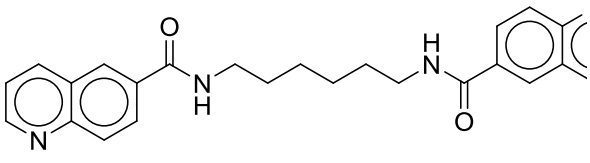
FADGUO



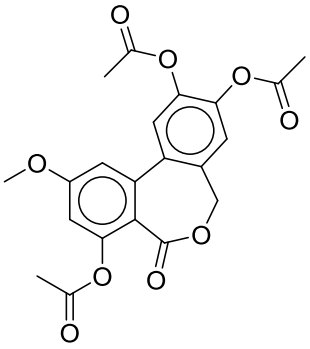
IRAVII



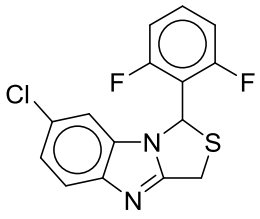
TAPPIK



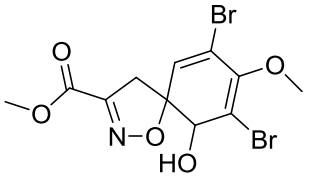
IFUZAM



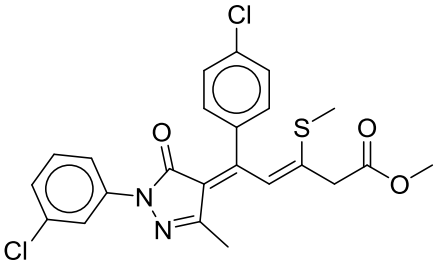
AXBZOP



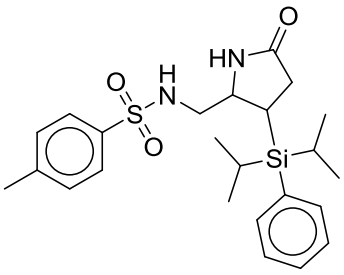
QIMBIX



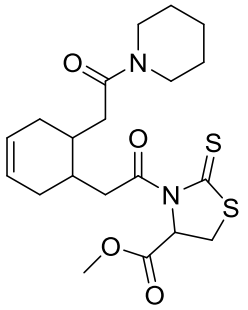
UKEMON



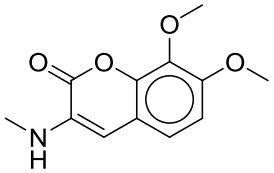
WAPGOK



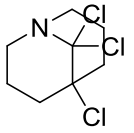
HUKDOH



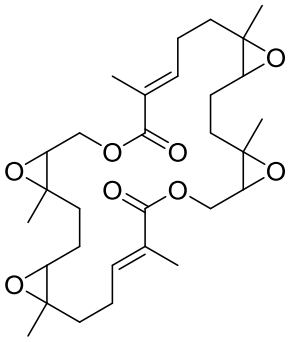
DIBREL



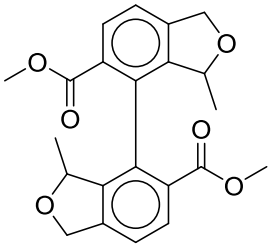
RULHUD



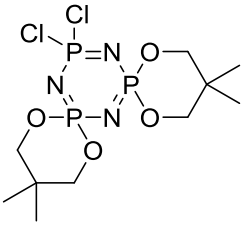
JAHYEW



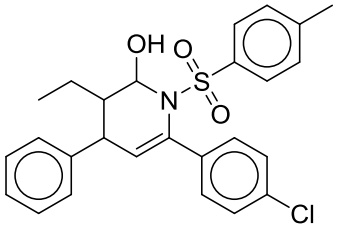
PIKZEO



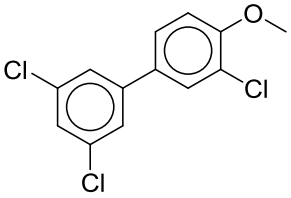
TIKWOB



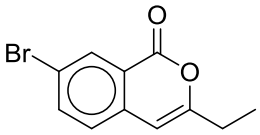
PEYMIQ09



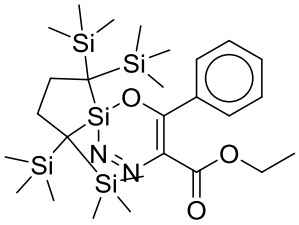
XOPPIC



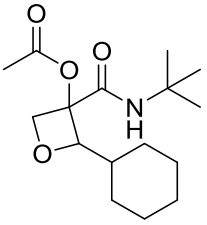
XOZJIF



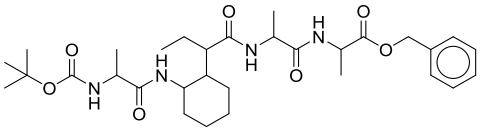
KUFGUP



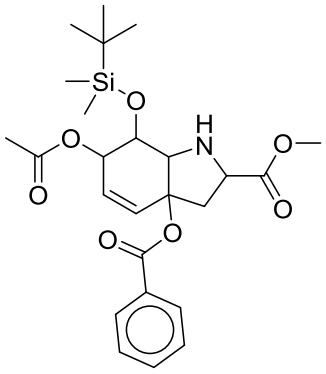
RUPVIJ



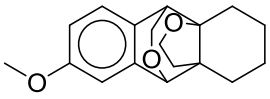
GEFBIE



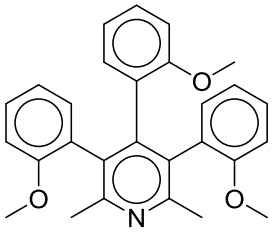
VEVMOA



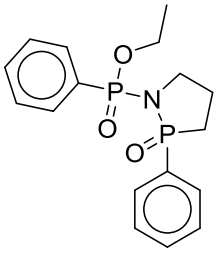
RAZWUK



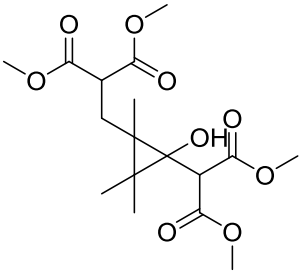
MANDEM



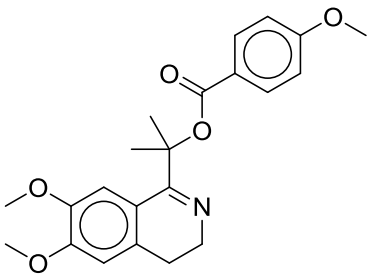
PIQTEQ



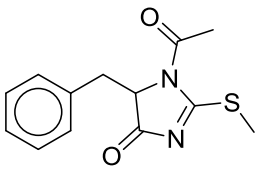
MAHSIZ



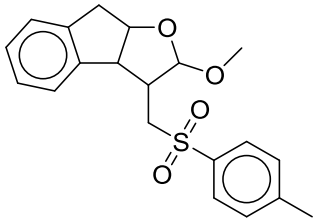
VOMSIZ



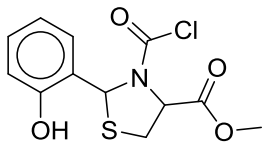
LAGLEK



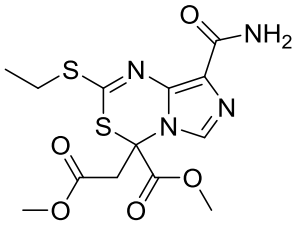
LETGEW



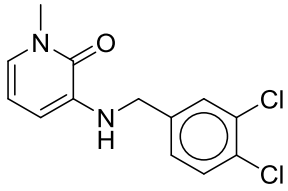
OLAHUE



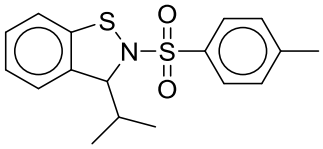
AXIMEZ



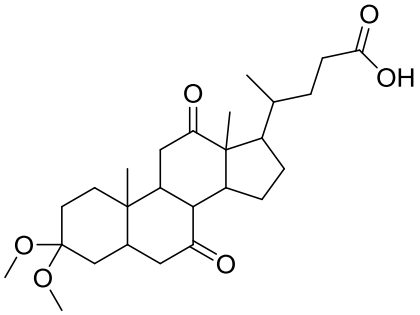
HUHLEC



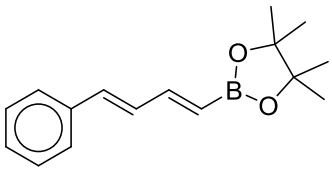
IDEGOQ



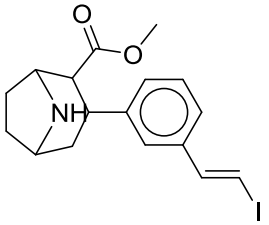
KUKZEW



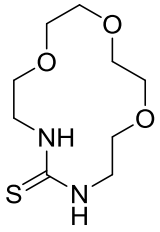
GOKDEQ



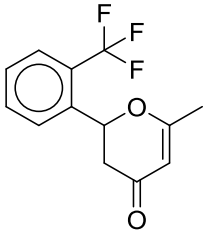
DAHYET



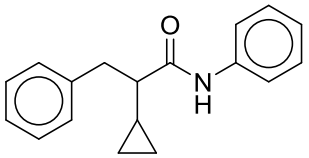
MIDKOA



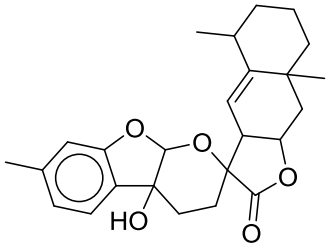
BAYSOJ



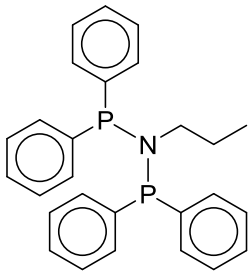
VAFFEP



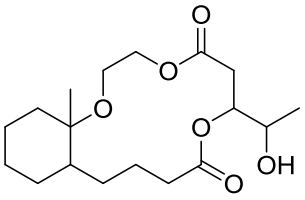
ZEGGEZ



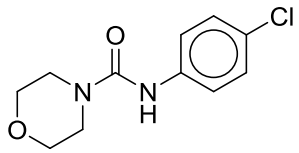
MEWDAT



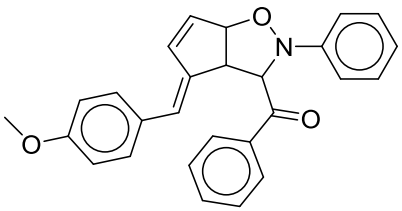
QEGWEG01



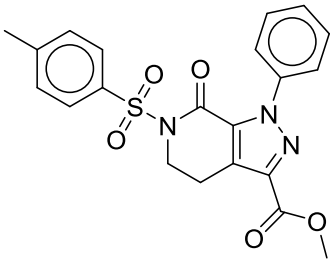
BANLEI



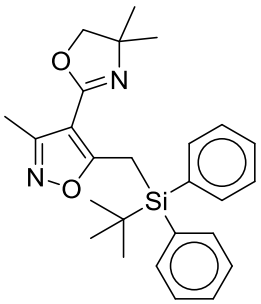
EXIMEE



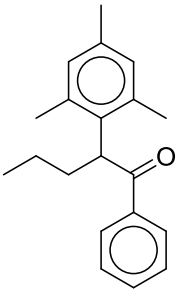
NUNMAK



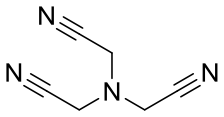
MEBTIY



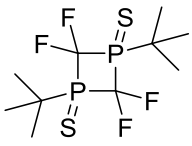
GIDXIA



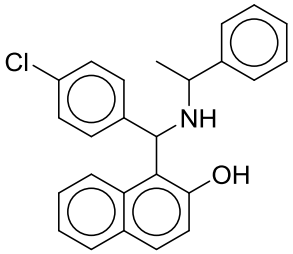
VOVHET



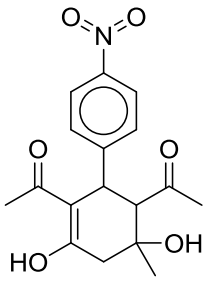
CIRWOR



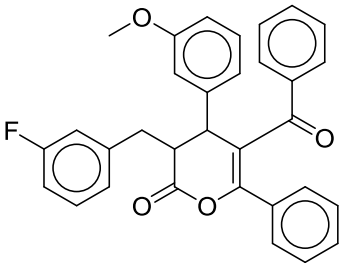
WILWET



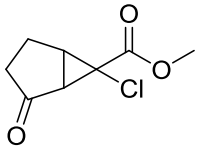
PARXOX



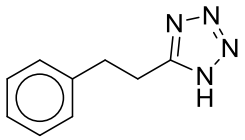
DAMPCX



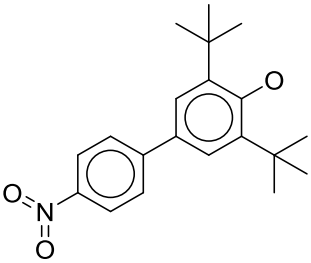
EZUKOA



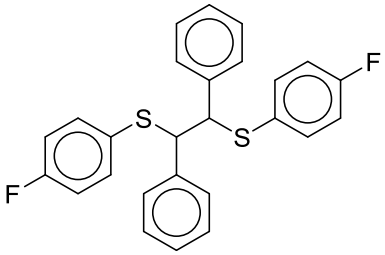
LAHKAH



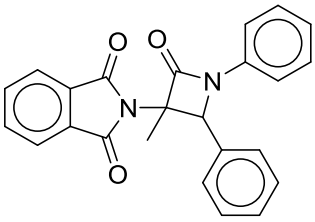
ELAJEI



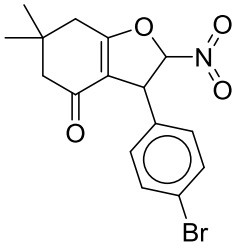
RUGHIM



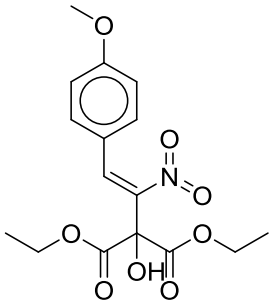
DUZREW



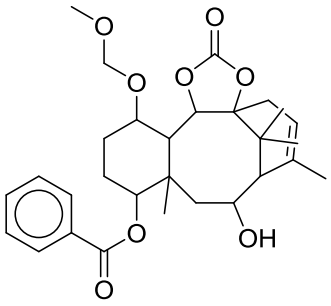
ZERJOV



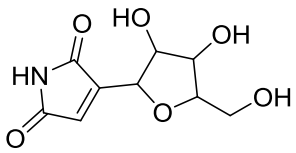
USOLUK



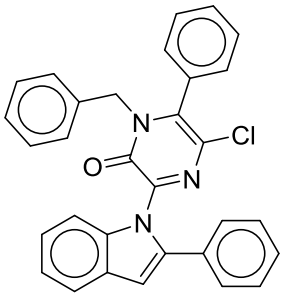
NUJFII



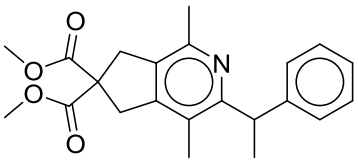
XULMUO



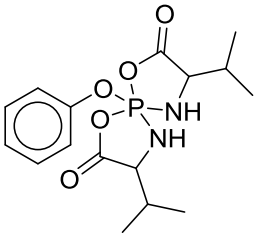
KELKOB



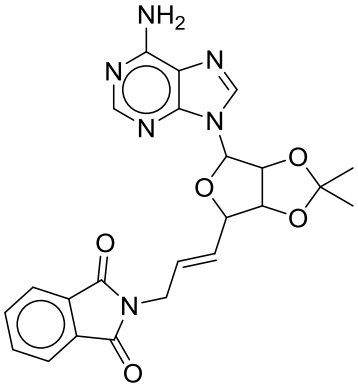
WIXWIJ



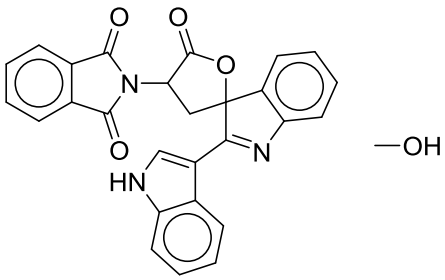
MEHXEE



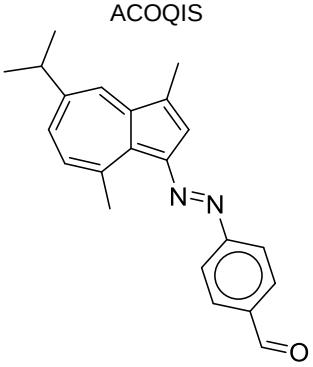
VELSEM



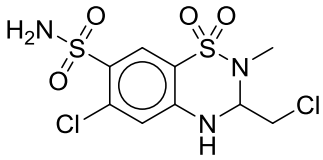
ACOQIS



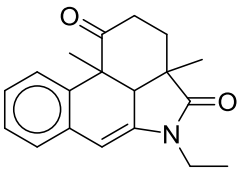
MEXPEM



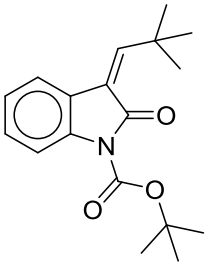
EFEQUD



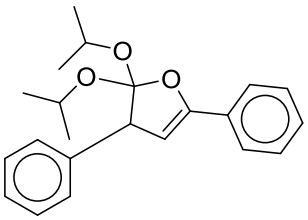
IFAFAX



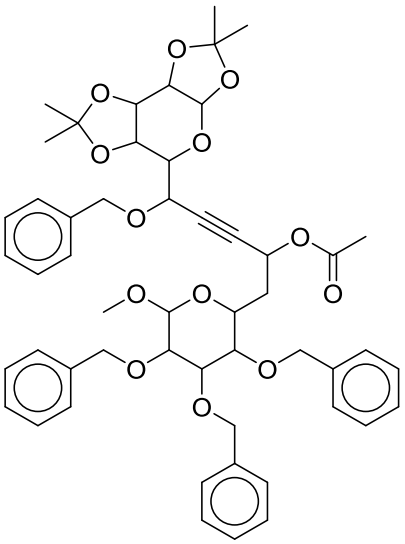
AEBDOD10



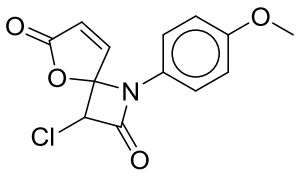
PEXKIO



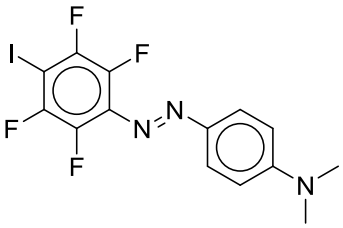
OQAXAG



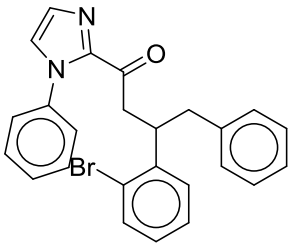
FARYOM



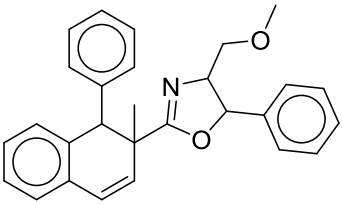
WATTEQ



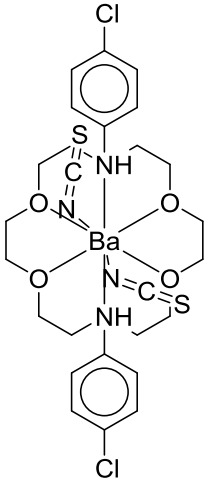
TEGGET



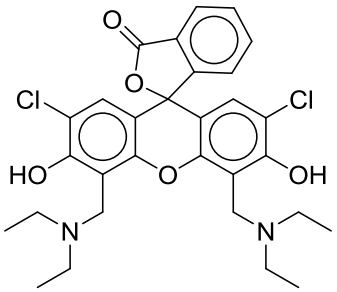
OREDAS



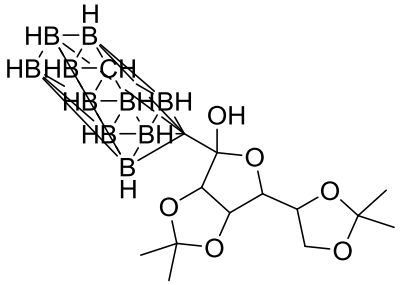
COVPIM



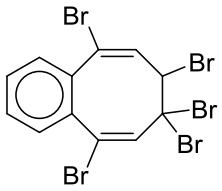
ZULHIX10



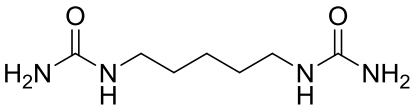
SUQVAC



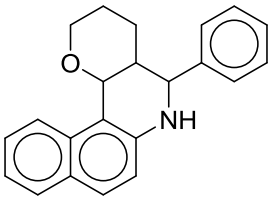
LAXWOW



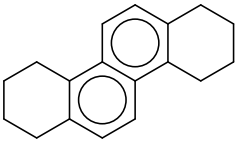
FADBOC



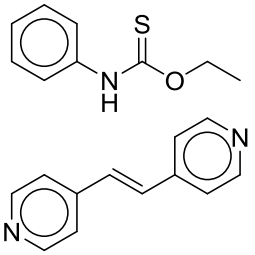
JAXKID



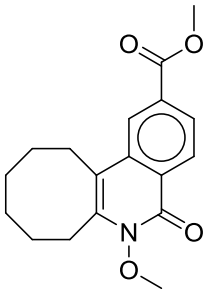
QUYVAI



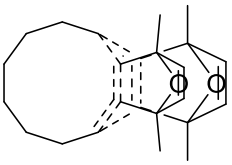
OHCHRY01



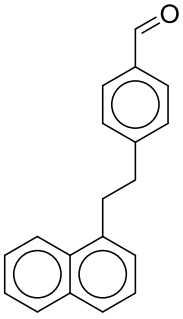
UHOSEQ



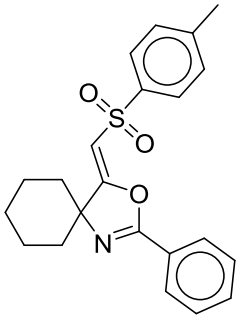
COWBIB



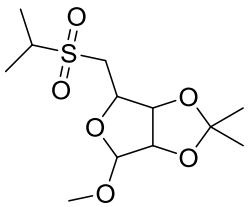
TULYEE



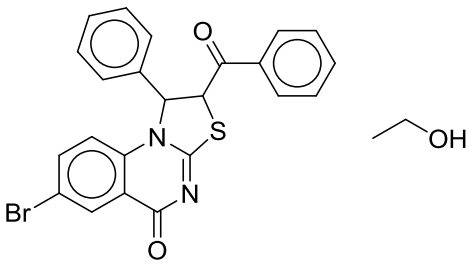
IJAMUD



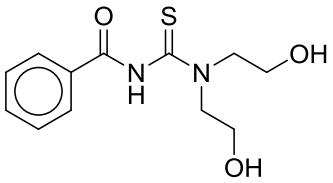
EVECOZ



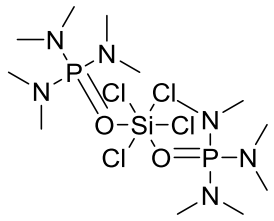
POHYAN



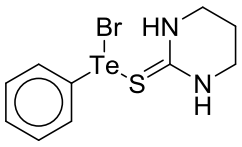
MAKPET



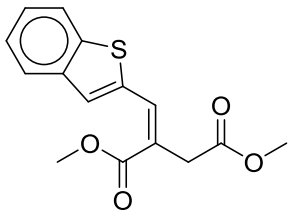
ZAJWAI01



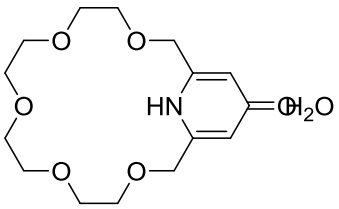
NITSOZ



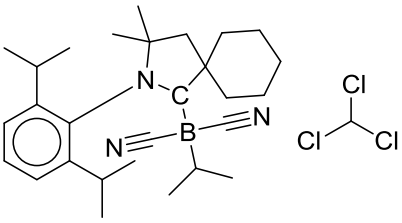
GANHOS



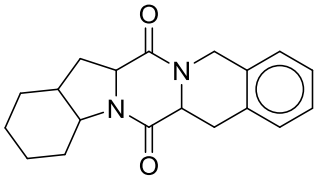
MAJVIF



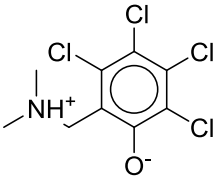
DABWUY10



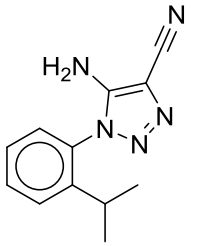
DIGWEX



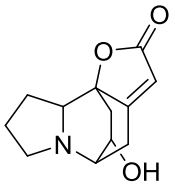
TUNZIN



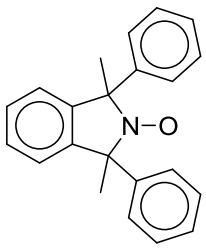
PIHZUB



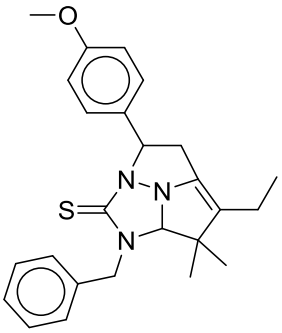
EMODIU



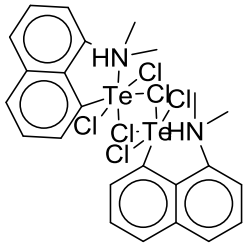
WECXUZ

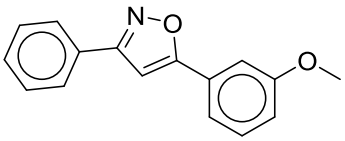


HEGROA

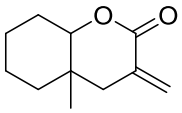


PIBRUP

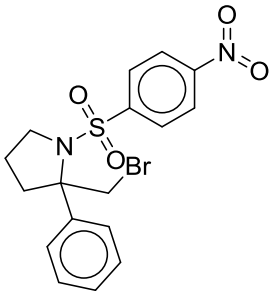




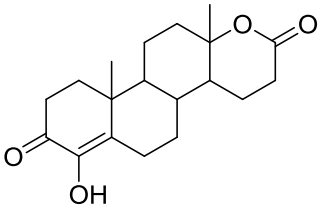
IPAWIG



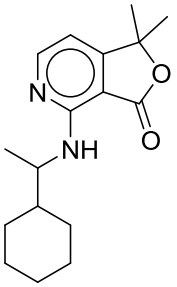
WIGMUW



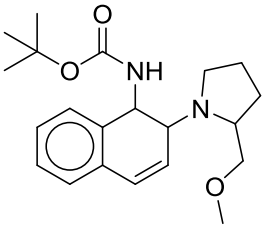
MAJWOK



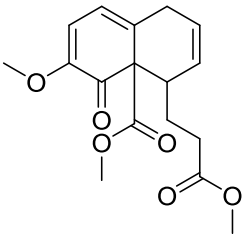
EVUSUK



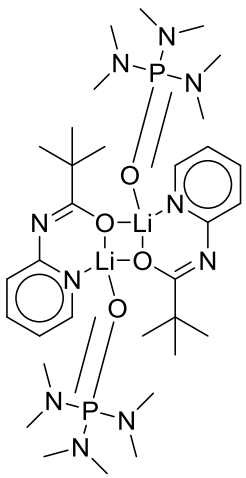
NAQWEH



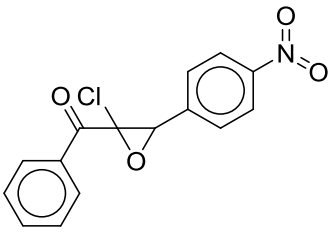
JACDID



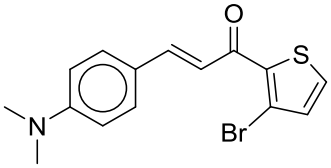
DOJWIJ



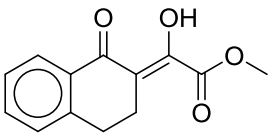
NAYLUU



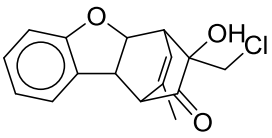
TOCWAJ



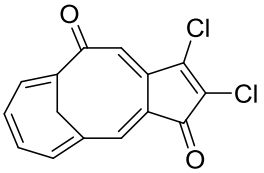
UHIDIA



DEXWAF

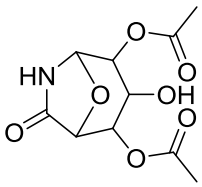


SATCIZ

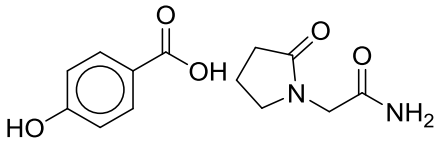


CUJLAW

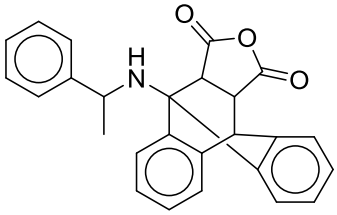
POJWUG



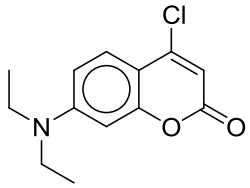
SINMIM



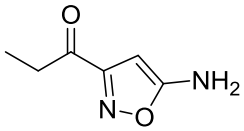
DAVPEW



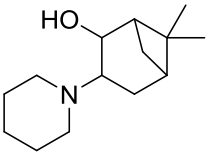
ODEMEQ



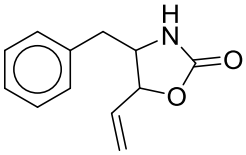
LELWAA



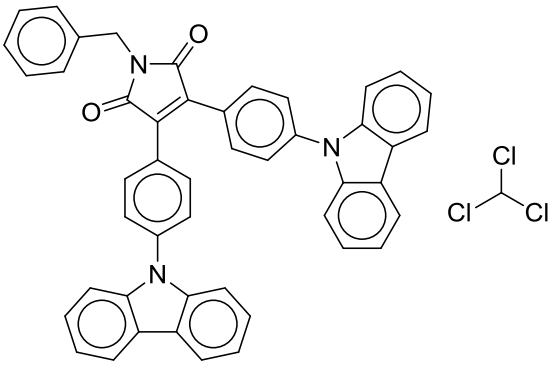
SAWMOU



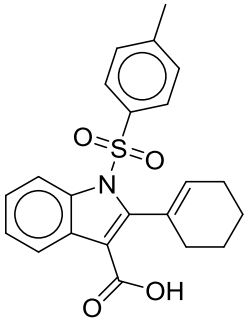
PORSIZ



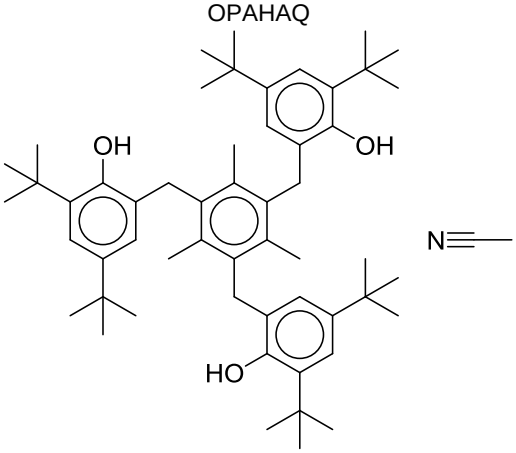
SAWDUQ



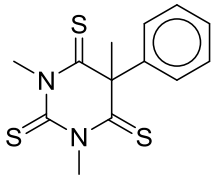
OPAHAQ



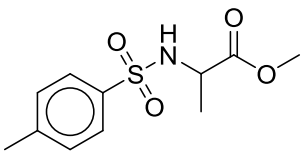
FEKMAL



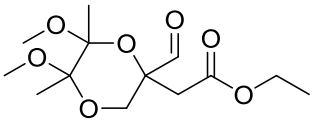
YUJCUC



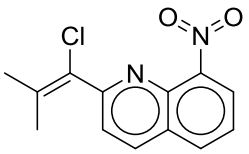
VUSLEB



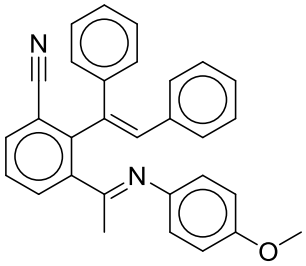
SOYPAY



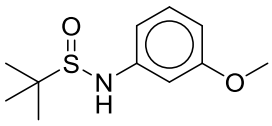
EGOTUQ



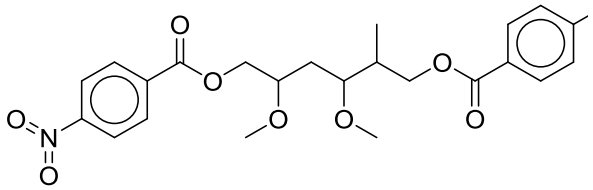
GIMQOJ



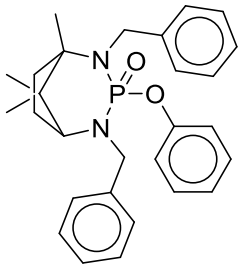
GUXKUH



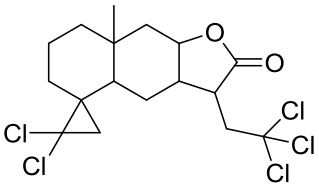
LAQVEG



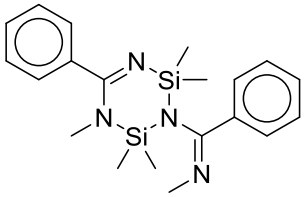
ISISUY



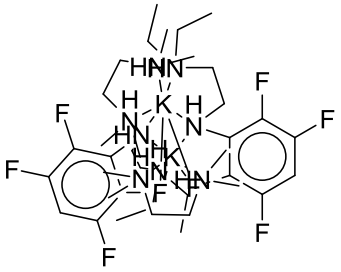
EMUQEI



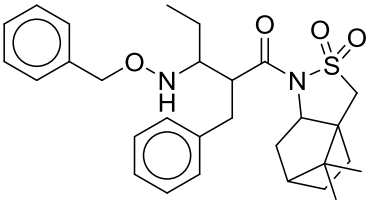
GEMGEM



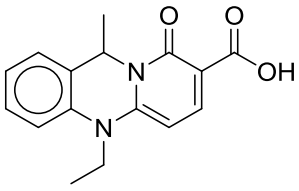
CIFREP



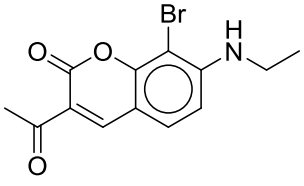
POSJEN



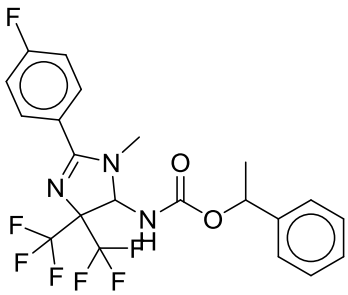
DAQNOY

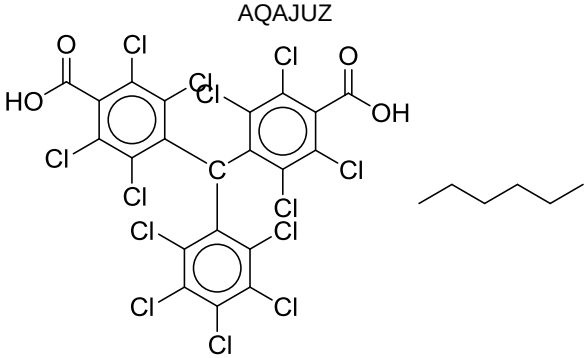


WOKNOZ

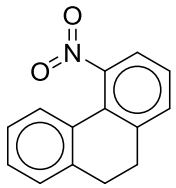


AQAJUZ

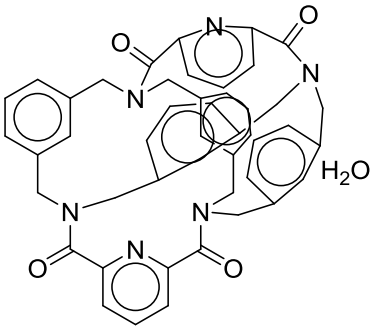




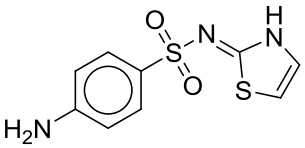
NEGCIL



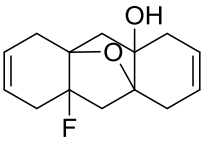
BELQOZ



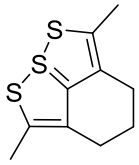
MEVZES



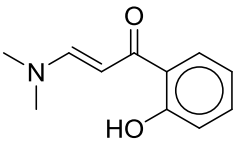
HUSSEU



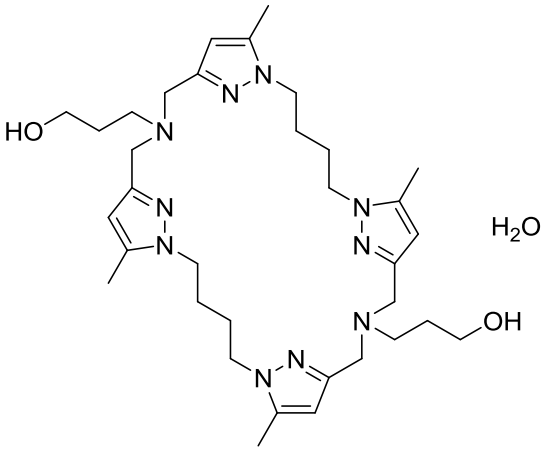
SUTHAZ07



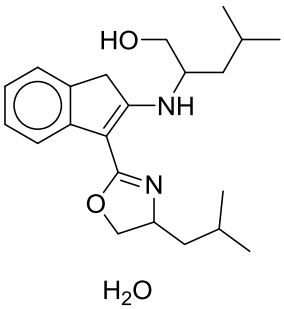
AXABEH



BILQEU01

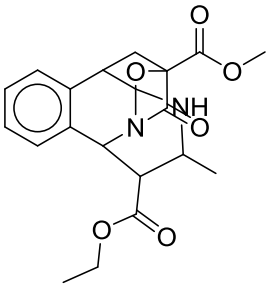


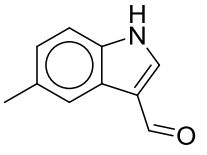
GUSRIV01



YUXJAE

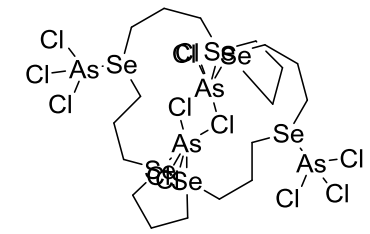
DOWHAZ



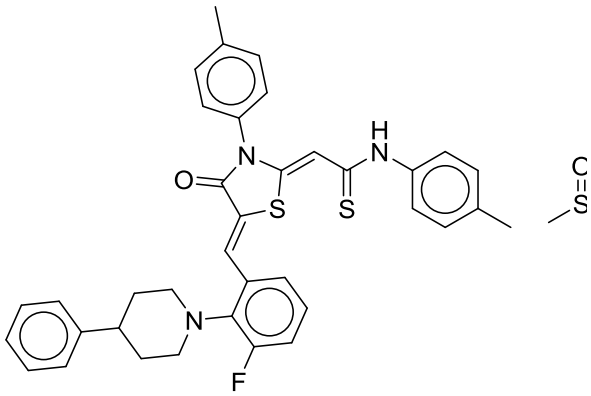


JUPMUC

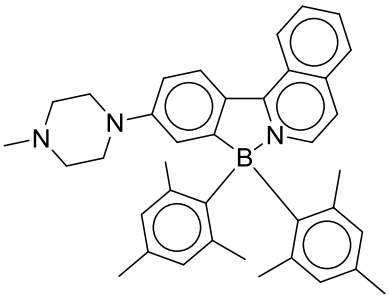
QECXAZ



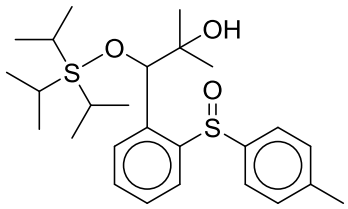
QUTLIA



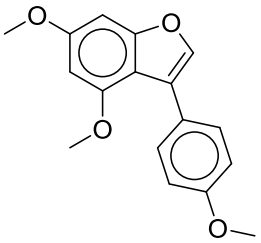
NUPYEE



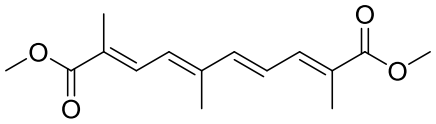
RUVXEN



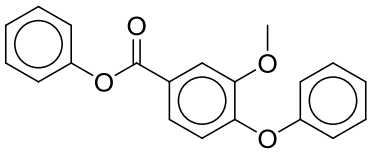
WIYNIB



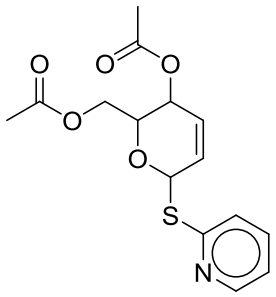
DATSEY



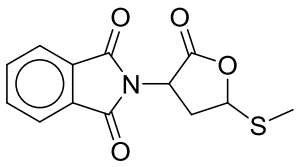
BUNVAH01



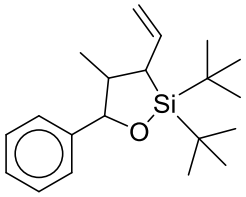
AYUWIB



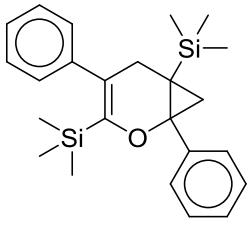
GAHTEO



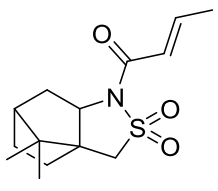
PLBULC



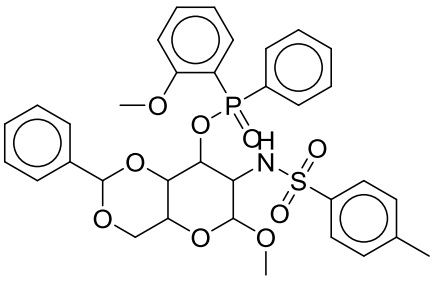
XUWMUZ



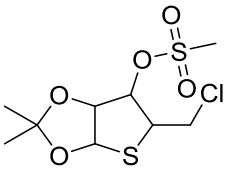
MAZNIK



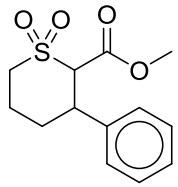
COYWUI



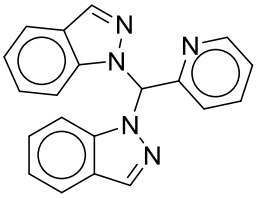
AJAKUT



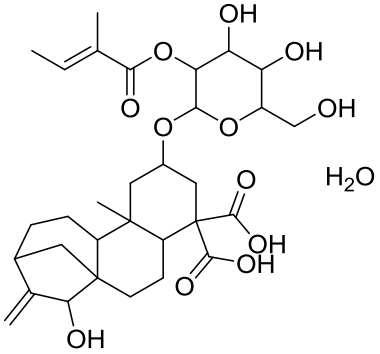
IMSTXP10



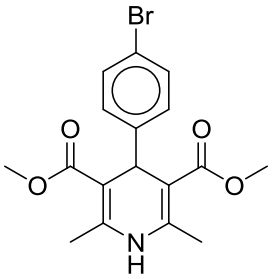
FUNZEU



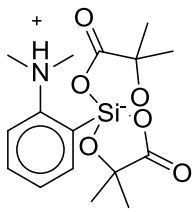
PEMFIW



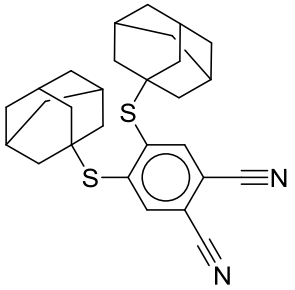
UCUYAT



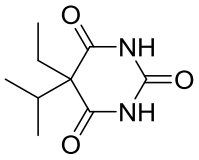
PIHTOR



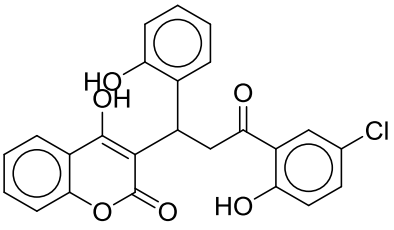
PIYTIA



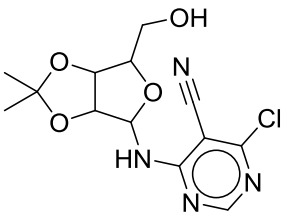
VOCVIU



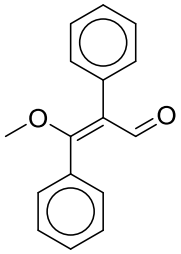
FUFTAC



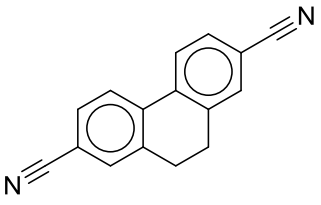
WULVOQ



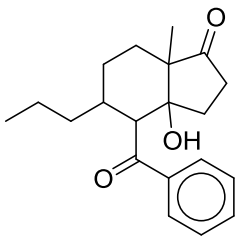
EMUYER



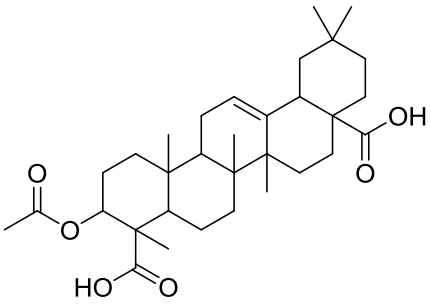
DEMTEU



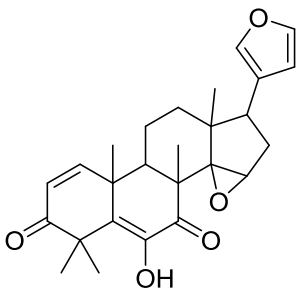
UWIDOU



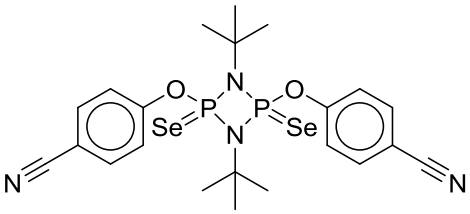
WUJVON



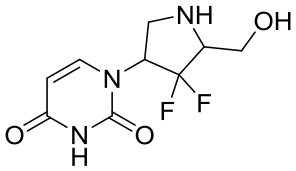
SUKKIS



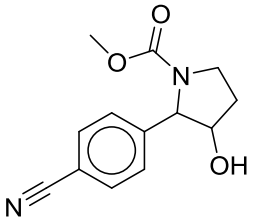
ZZZJFA01



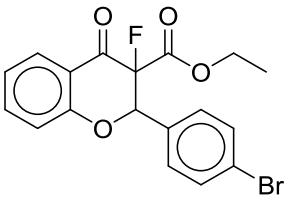
BULQIK



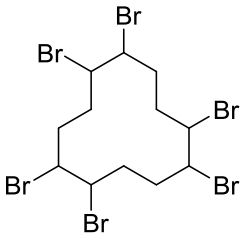
YOFMUC



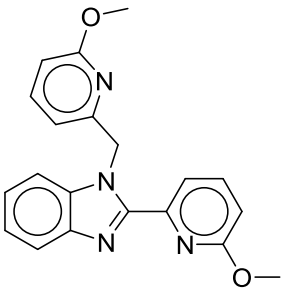
QAJBIN



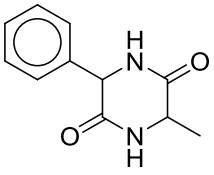
SUHBUT



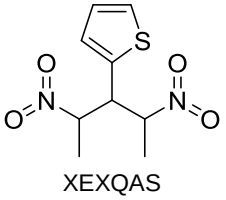
JATBOW02



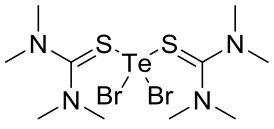
SODZIW

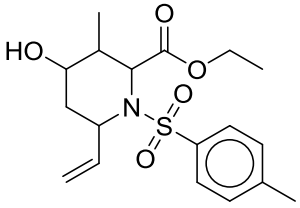


LEKKAN

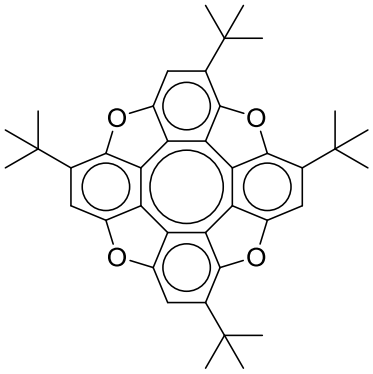


XEXQAS

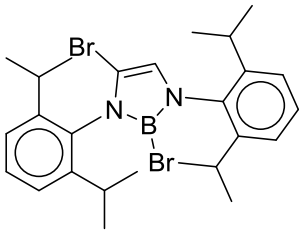




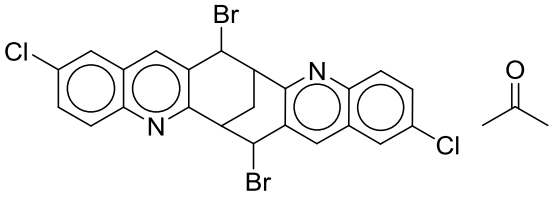
ZZZEIY01



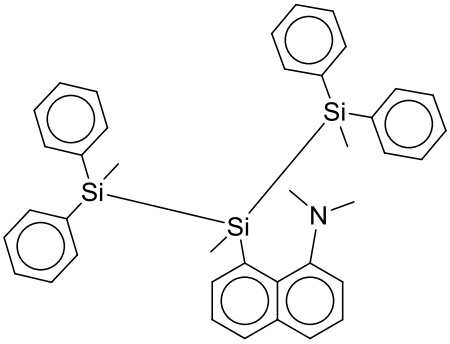
TIMGOL



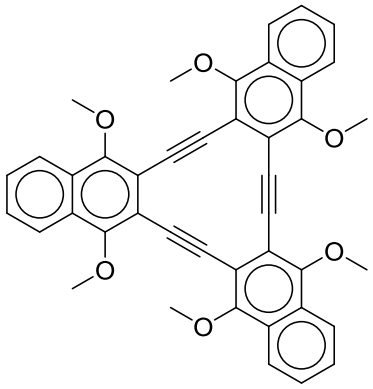
DAFMEE



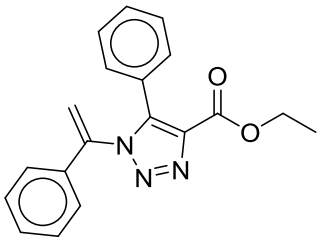
FUXQIA



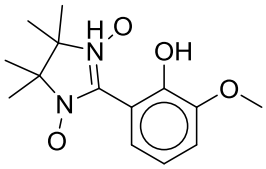
QETYOE



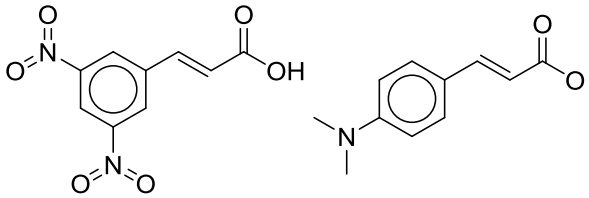
JUZTUT



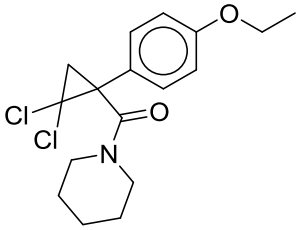
GIBREO



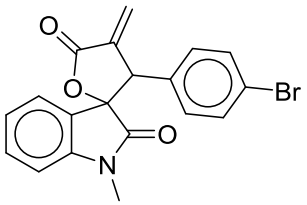
PAJWUV



VUFSEV

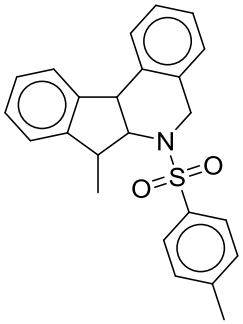


PICFAI

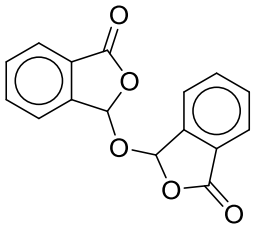


NERTOU

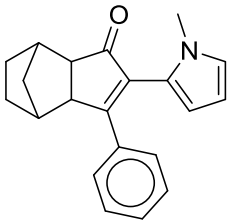
XIPQUK



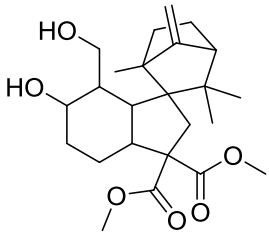
BORLIF



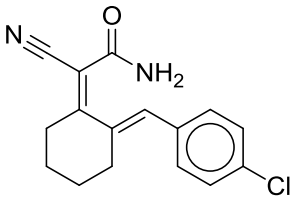
NOSYOJ01



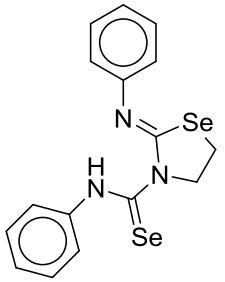
DADBIV



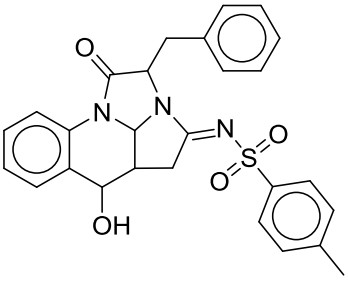
NAHYOK



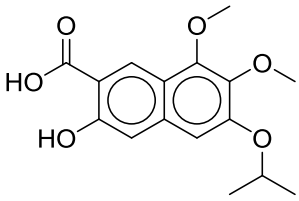
OLIBIV



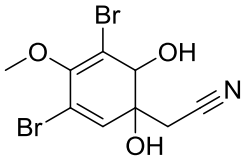
DATCAD



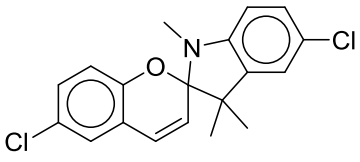
FIGFAE



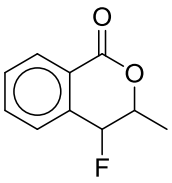
HETJEV



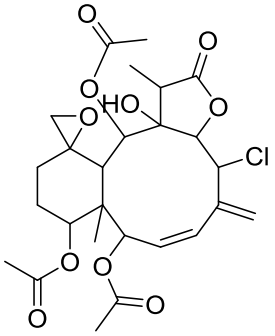
ARPLYS



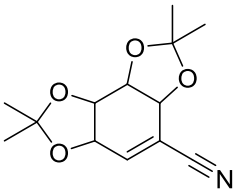
HODREY



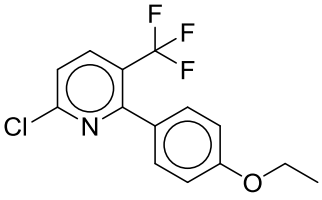
HAFKEH



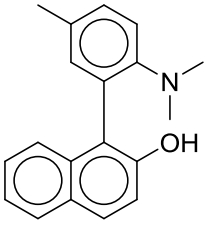
TISRUJ



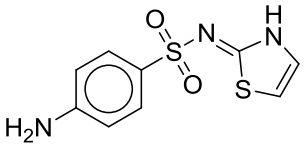
NAHMIS



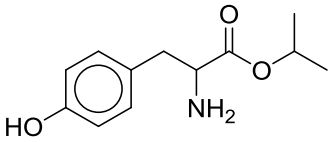
KEZPUC



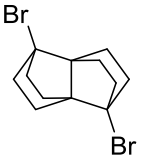
TAMNUS



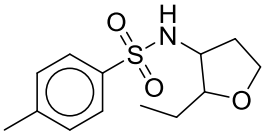
SUTHAZ37



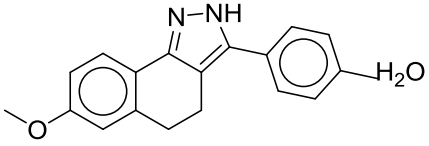
SECTAX



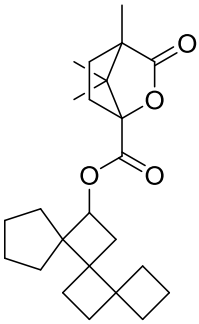
BAHZEP01



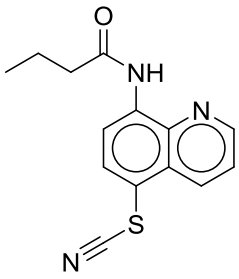
MUJBIE



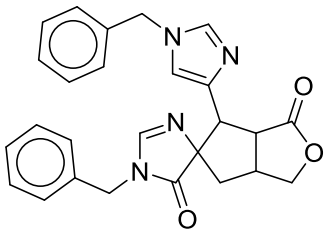
JIVMAD



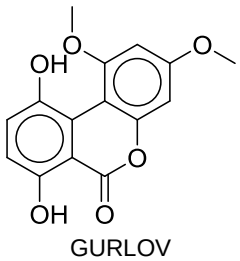
LITPIO



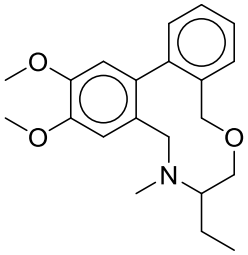
EZEGEX



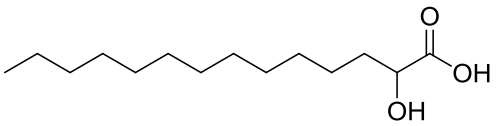
NUNHEK



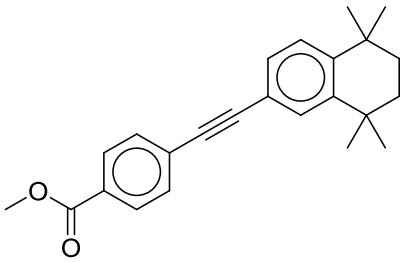
GURLOV



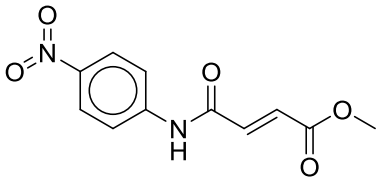
MELXAC



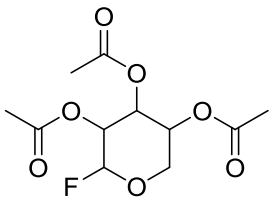
DLHTDA10



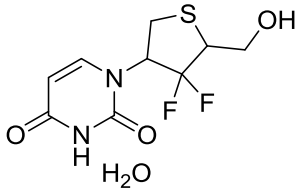
KOHHIZ



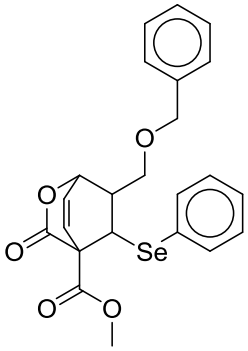
QAJWUV



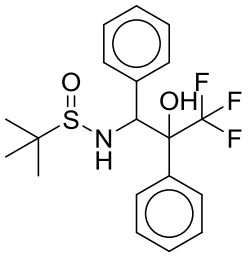
BIVTAR



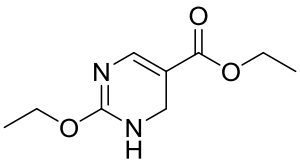
JEYKOO



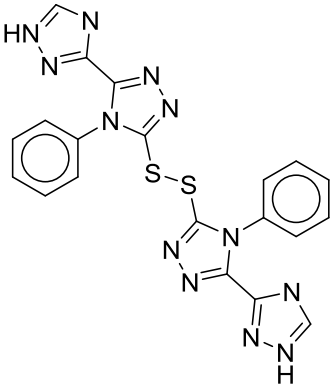
CIDPOW



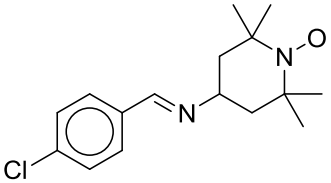
WOBNEI



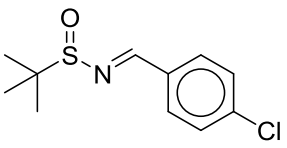
IDUZUD



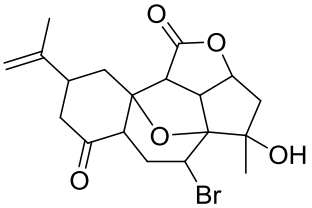
PAMTED



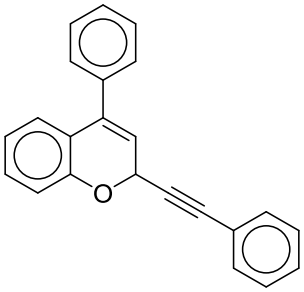
TIYJEQ01



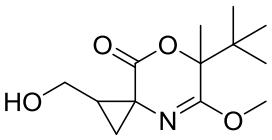
WICPOP



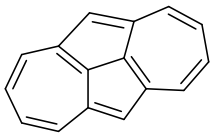
UVUXOA



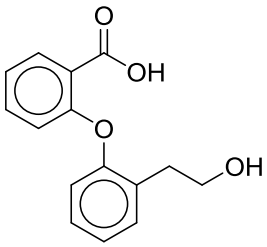
IGUPEH



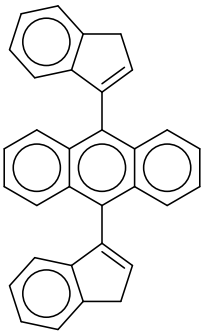
GUYNIX



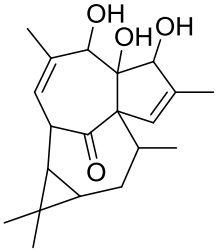
COLDEM



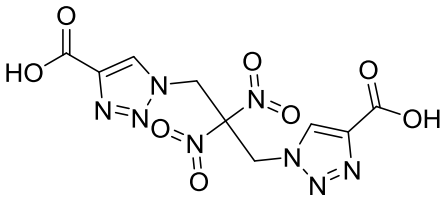
RESROW



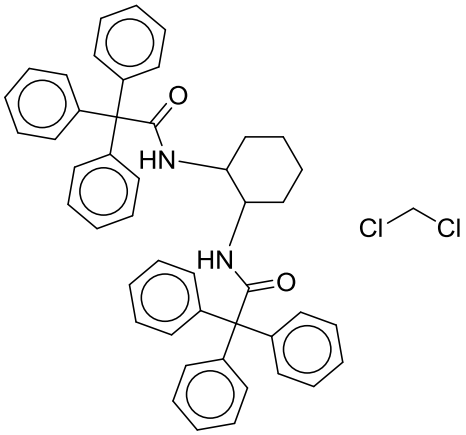
APOWIM



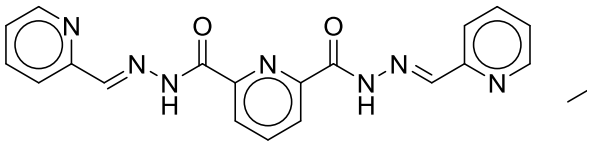
FEMPAP01



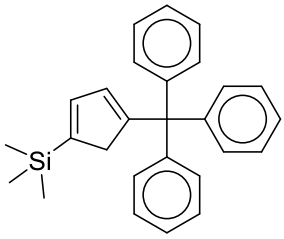
ITINEE



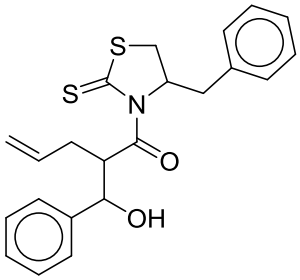
TUWKED



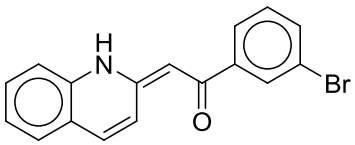
VUZREO



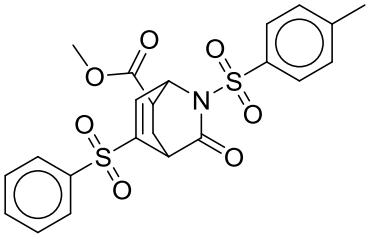
YIVXIK



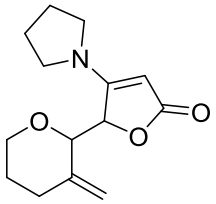
MODMAT



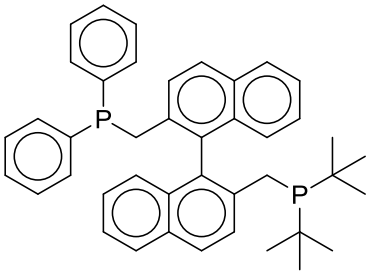
LOBHOZ



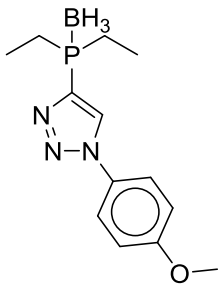
POBVIM



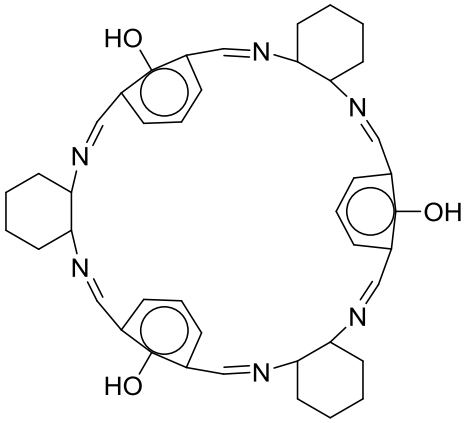
ATOYUD



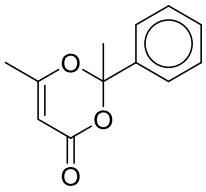
MOLZES



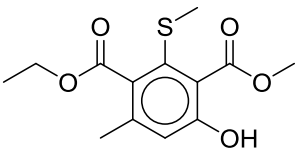
CODQAP



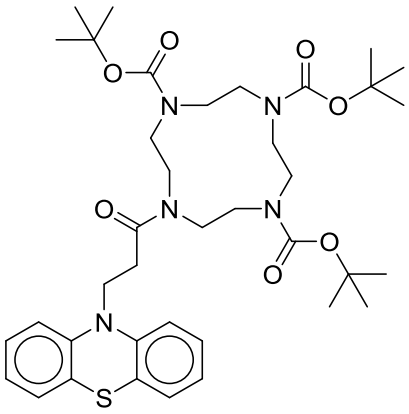
BUFTIG



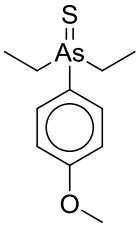
ZUSYIV



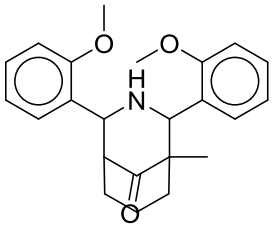
GILHUG



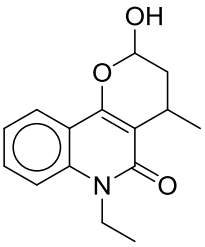
OBUSEJ



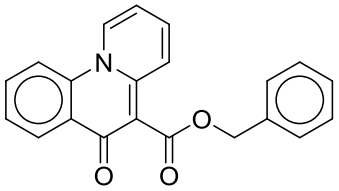
CINXAY



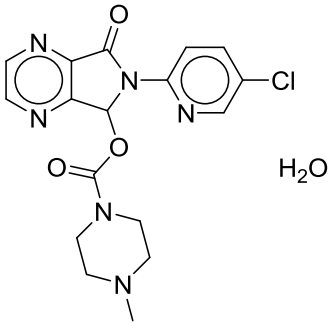
RUGPOZ



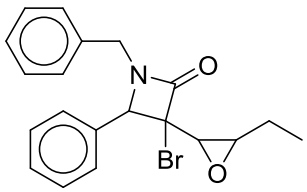
GOZPEQ



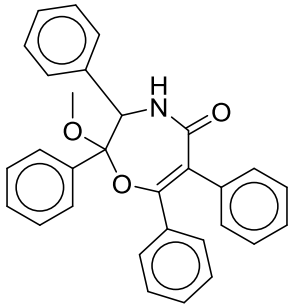
ZUYKUB



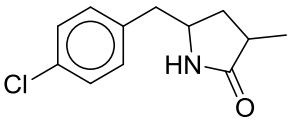
UCUVET



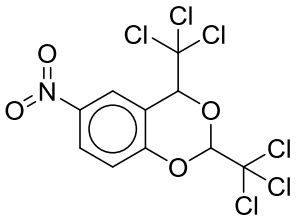
VEWMAM



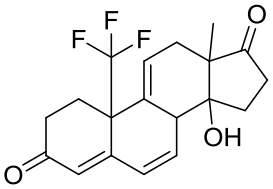
SUSDER



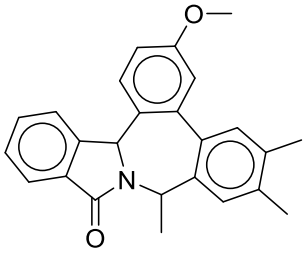
JAFCOI



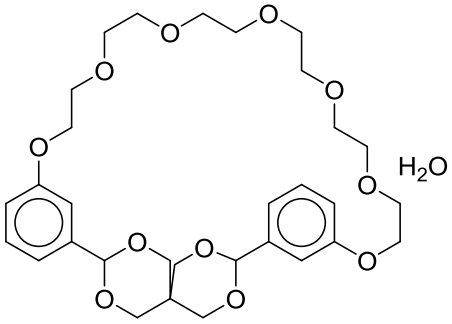
FIFXIB01



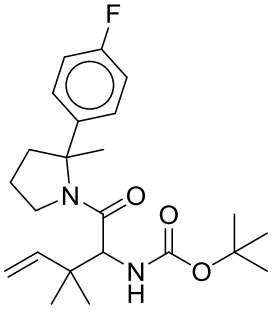
NATMEA



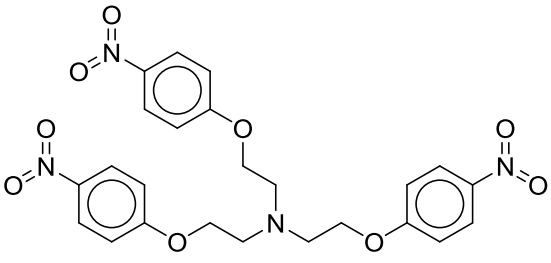
HESZAH



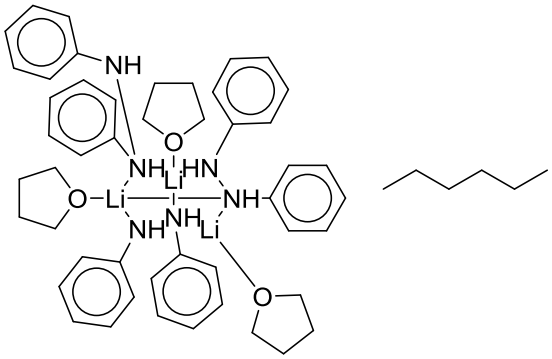
EKONIB



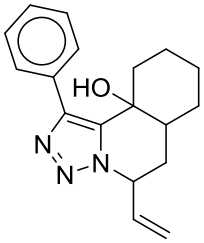
DAHQOV



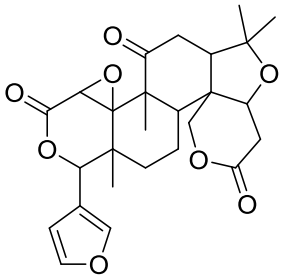
PURSEB01



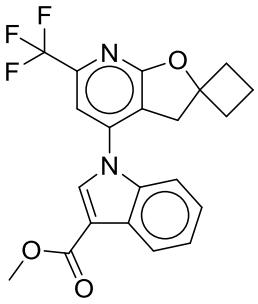
PUFNIP



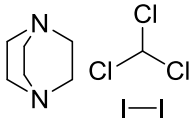
RIZFEN



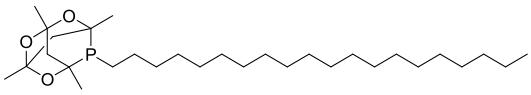
ZZZSSO03



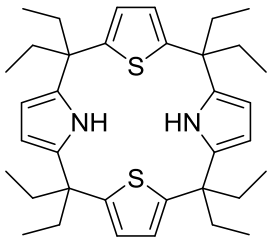
UKOMAK



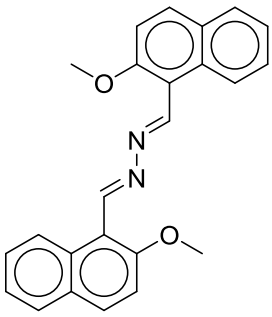
HEKZII



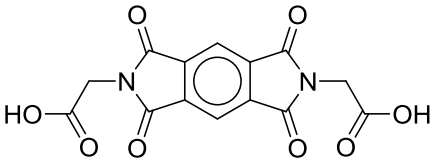
VOFYEV



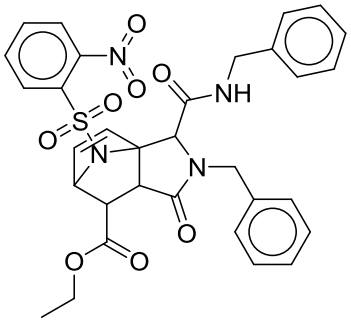
LOHVEJ



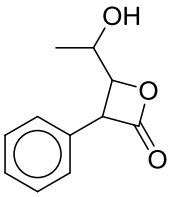
TAZXOI



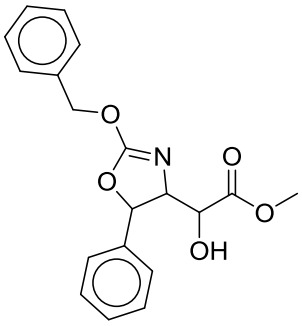
GENMOC



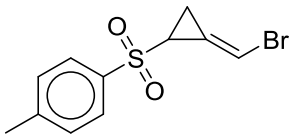
EVATOK



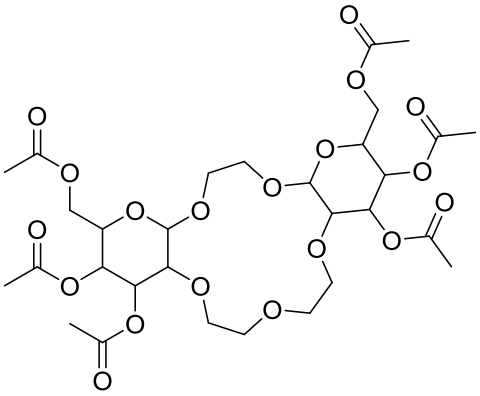
GOHCEN



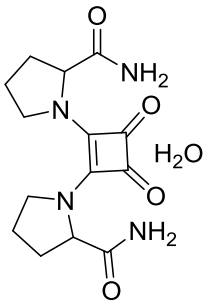
KEQPEB



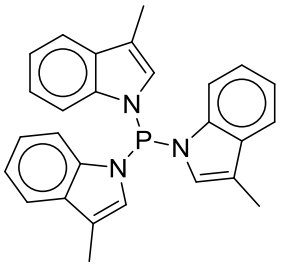
REDDOU



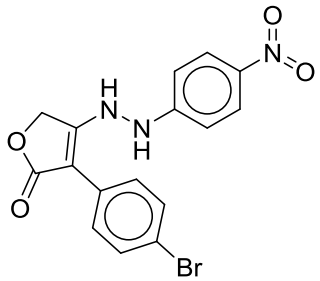
SOCJID



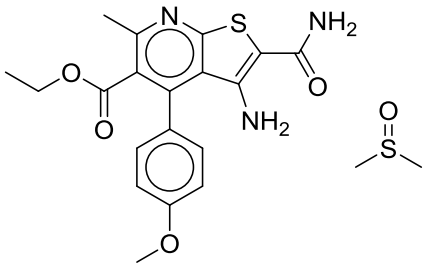
QORQIY



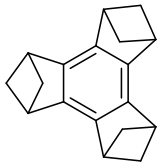
BIJKOV



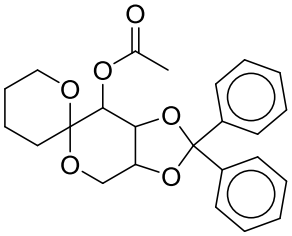
YAHSAD



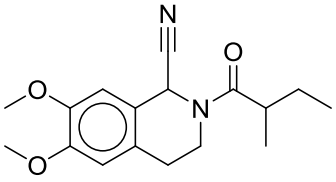
AWETIH



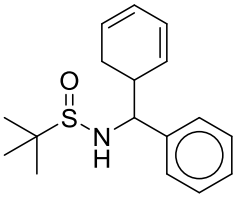
ZAKMIH02



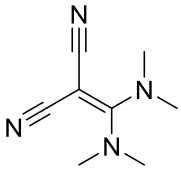
TERDUQ



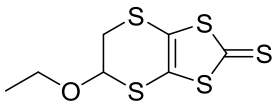
MIHZOT



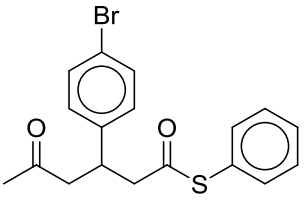
PODHEW



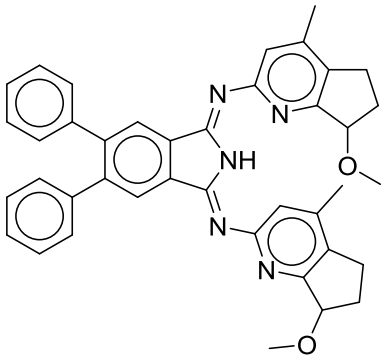
BUFGOY



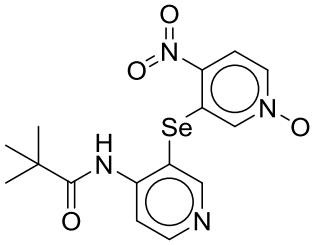
IRERUS



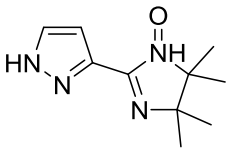
LUMSET



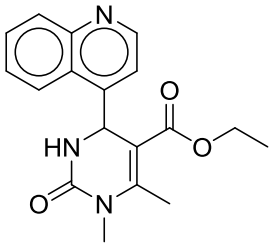
VENROX



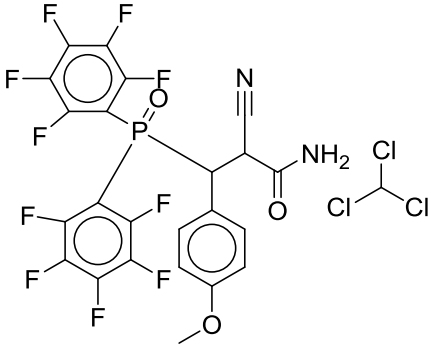
YADVOP



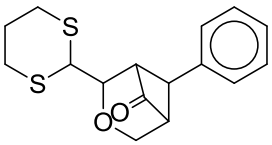
WEHKOK



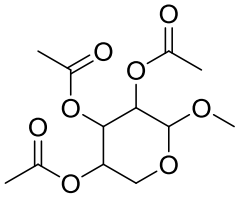
DUZHAI



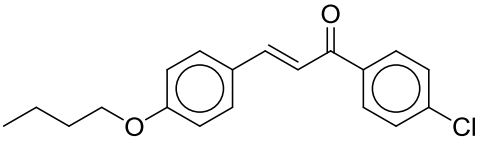
AYAROI



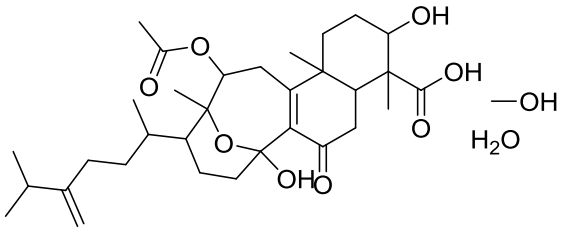
SARPUY



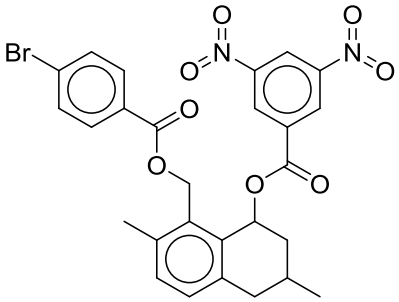
LYXPYR



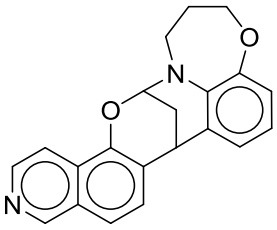
TERTAM



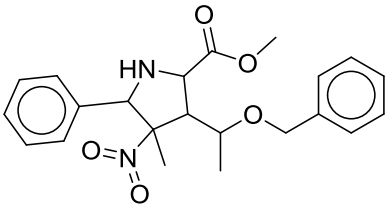
MOXXII



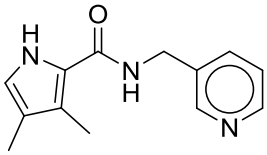
GICVOE



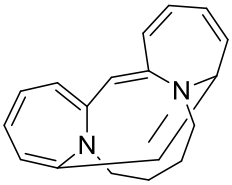
HACHEA



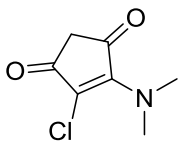
QURVOO



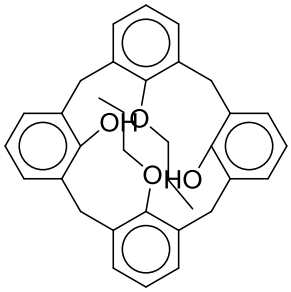
AMEWAS



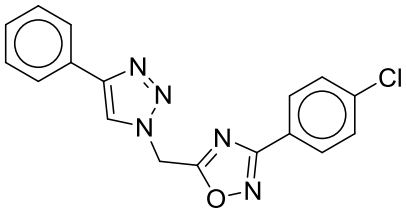
DINGAI01



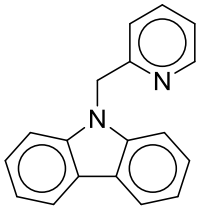
ROBYIQ



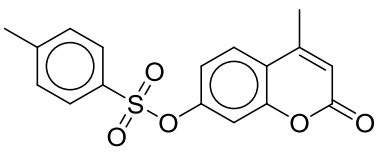
PAWTOW01



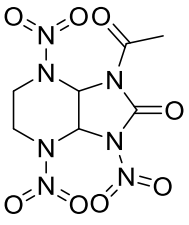
HAKHEJ



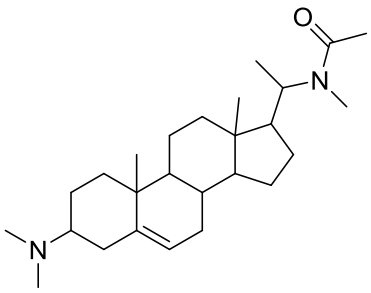
HAYWOU



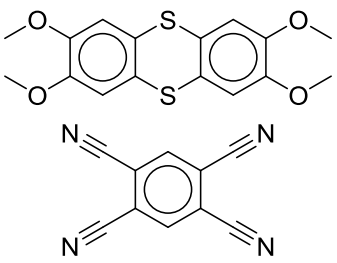
YAXTUO



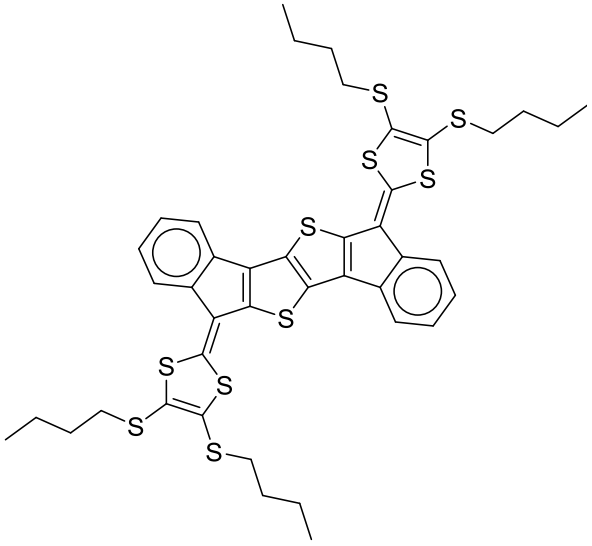
JOTDIF



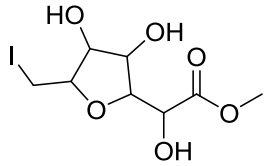
YAGMIE



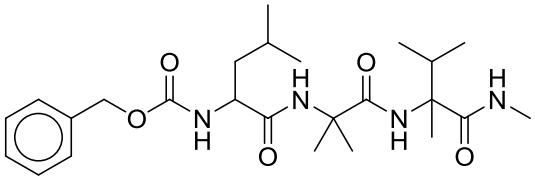
RIKYUF



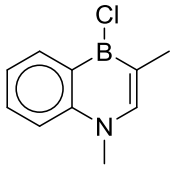
MUMMIS



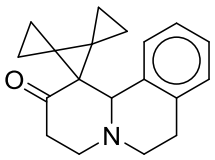
QOWQEY



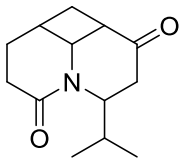
LIXQOZ



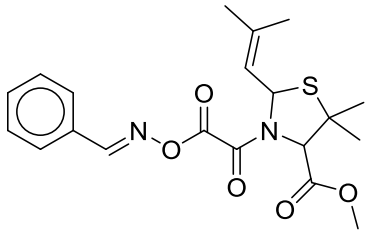
DAHYAP



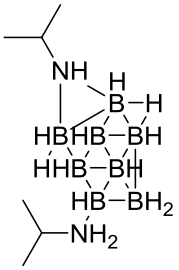
UHEYOV



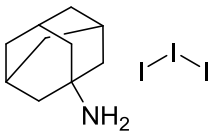
WOYLUR



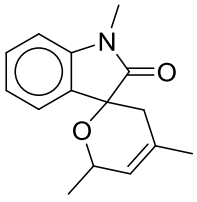
ETOZES



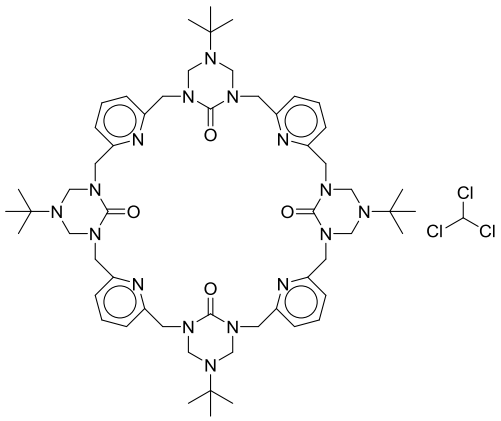
NIHBEL



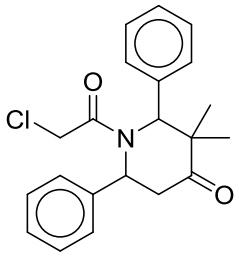
MITJUW



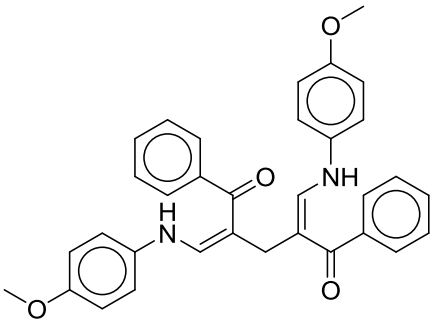
WUKRAW



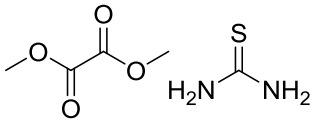
MAMNIZ



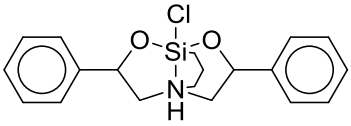
ROGGOK



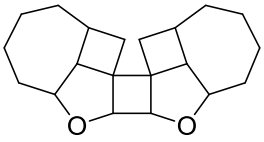
FUYLIV



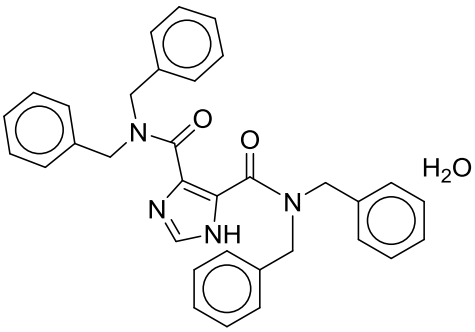
PEMBAL



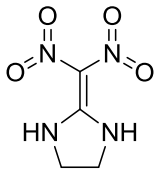
JEHBEE



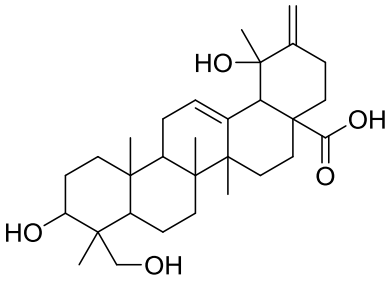
SEMKOK



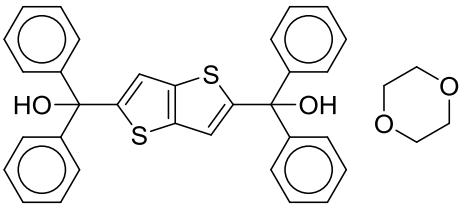
HUKLEE



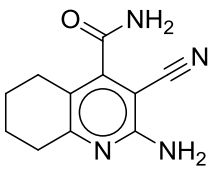
KOVHAE01



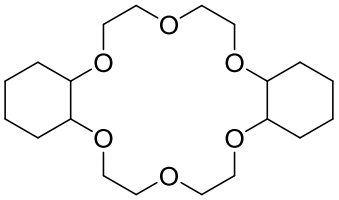
ELAKEH

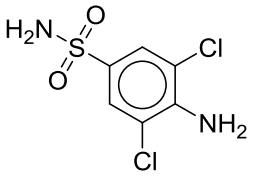


LUJZEW

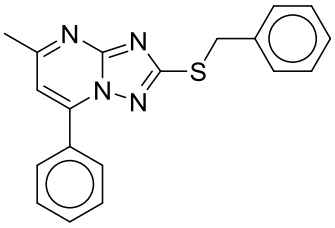


RENTOT

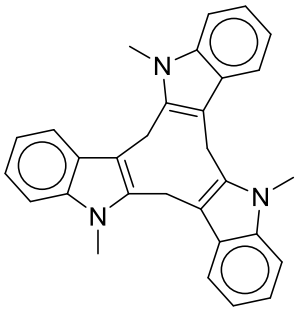




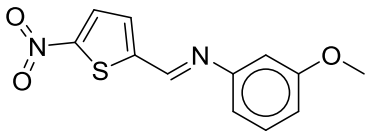
KOGSEE



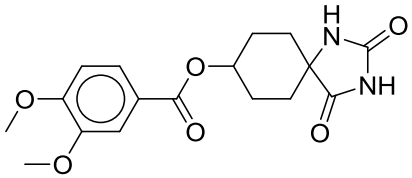
KEFRIW



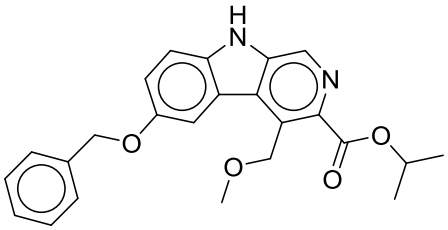
GUGJIC



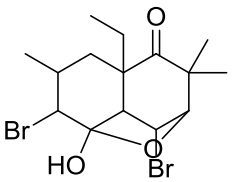
QECTAV



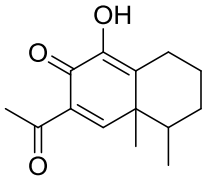
ZASYEZ



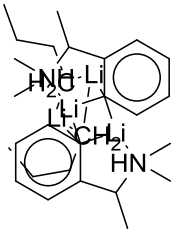
YERRUI02



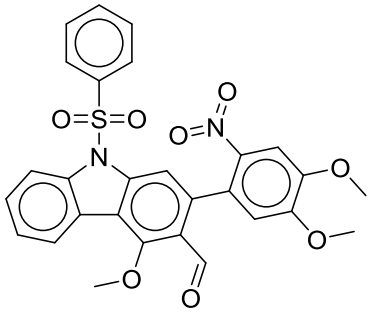
ZELNAF



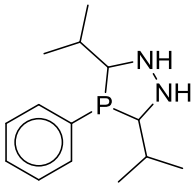
KUHTEO



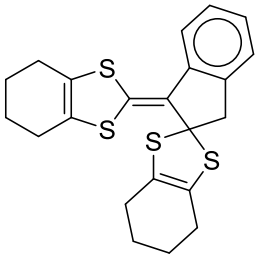
FESGUG



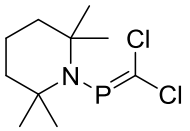
WOCMUY01



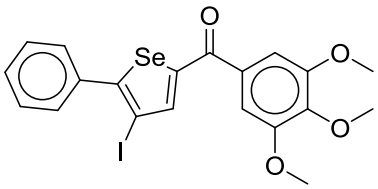
BUHQEA



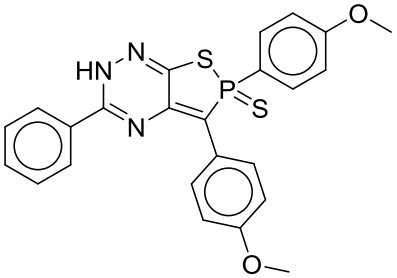
JUYGAL



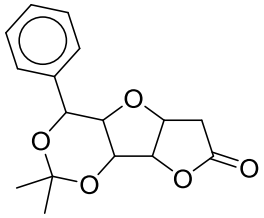
SOVPUO



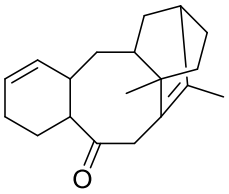
SAGTOM



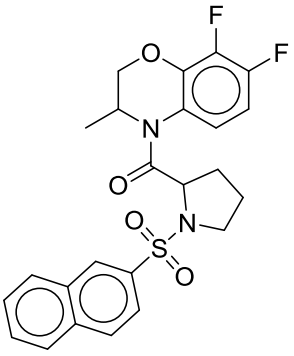
VEKXUE



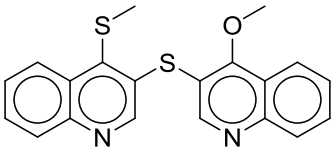
TOLJEL



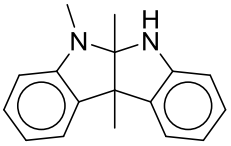
JILKOE01



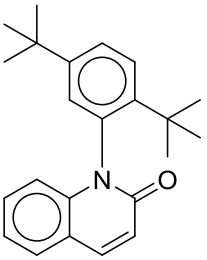
JOLPEH



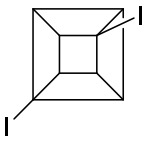
YICGIA



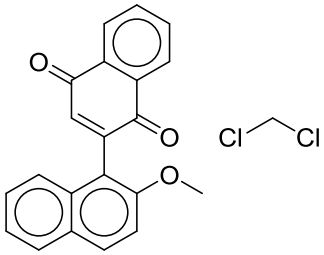
UNIDIF



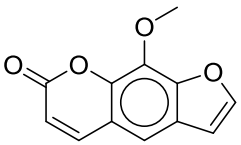
NEFPOF



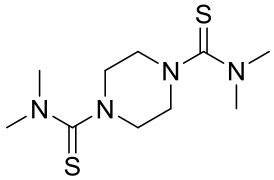
GIHTIA



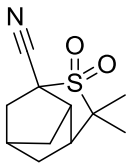
RUJPUJ



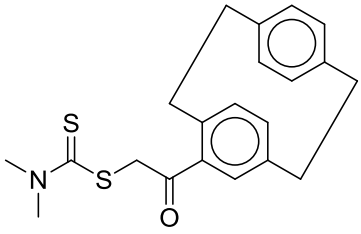
XANTOX



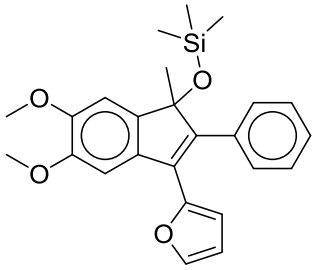
VIPPAM



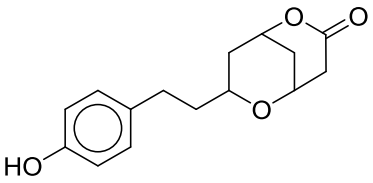
KARYAD



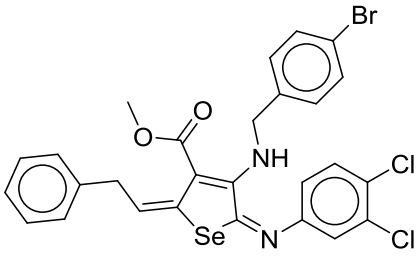
ESUHOR



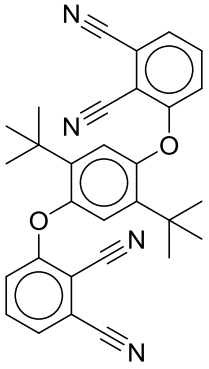
XOSWAF



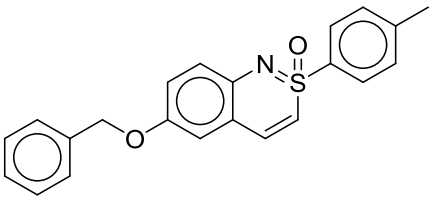
BUSNUA



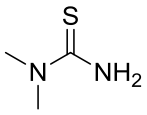
IGATIU



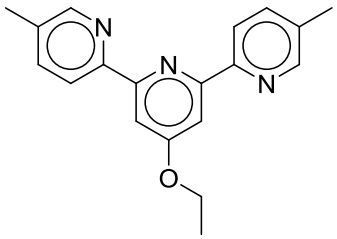
POYDUD



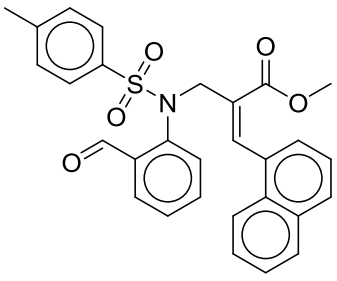
BOCJAE



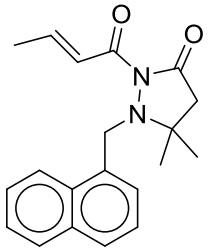
WIFKOL01



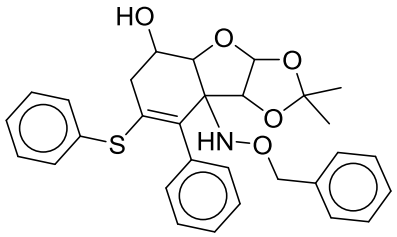
HOTHED



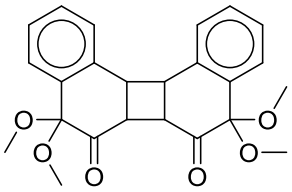
GAPTOI



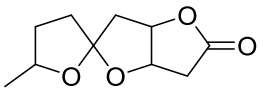
SEVYIC



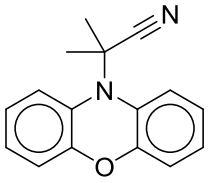
FIBBEY



JIQZIT

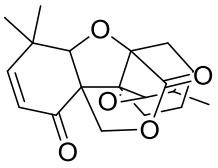


DAJSIQ01

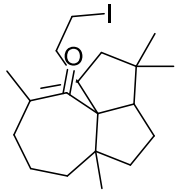


NOZXUV

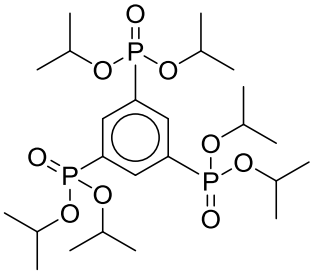
OGERAV



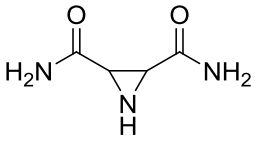
FIZNOS01



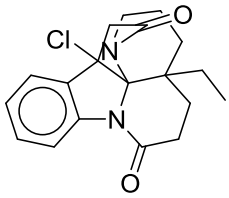
FOLKUM



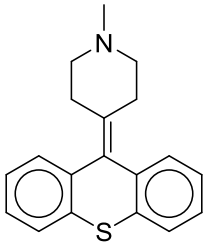
AZIHEW



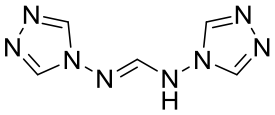
EREREY



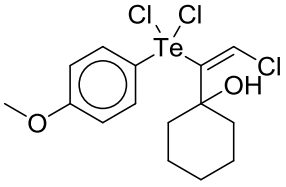
ONOTAO



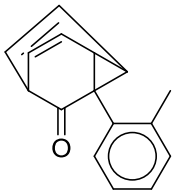
OQEQOS



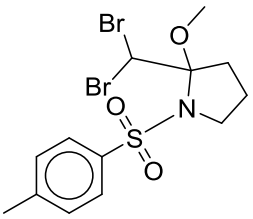
HAGLUZ



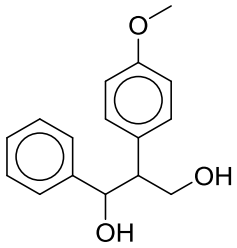
ASEHUB01



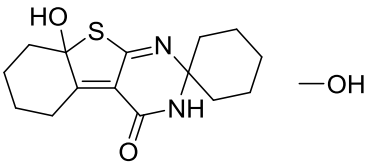
OTOGOV



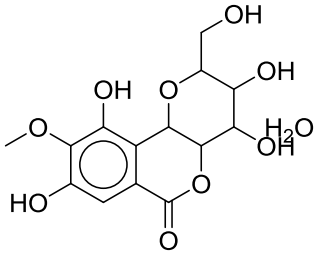
POSBIK



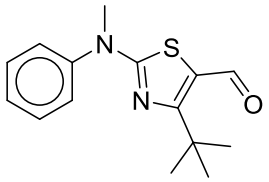
FUTMUC



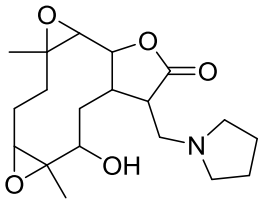
ELOPEC



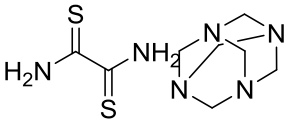
BIBGUQ



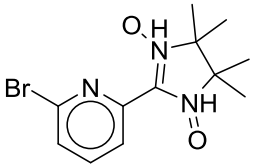
BUSPEK



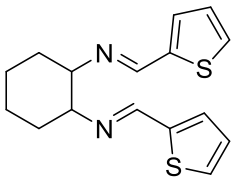
EVICUI



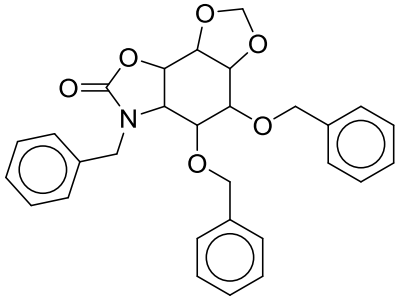
DOKHAN



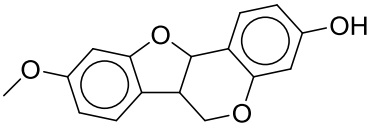
RAYTEQ



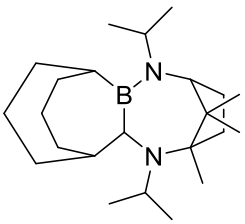
BAZNEW

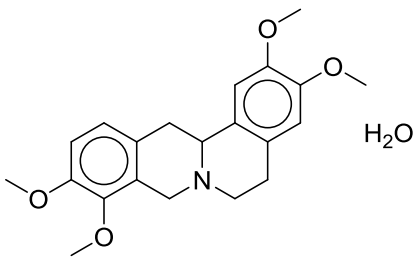


WALFEV

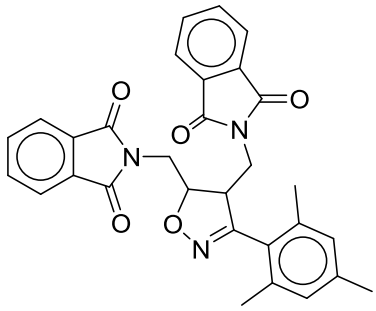


CUCWIH

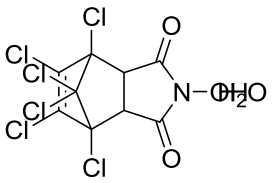




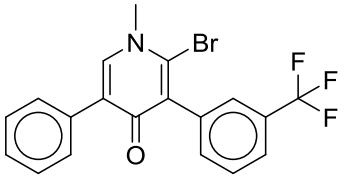
IQIQOQ



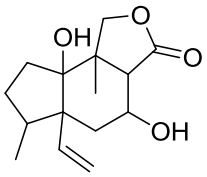
SEDJEQ



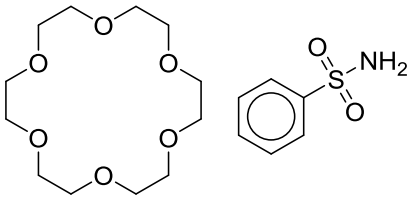
ZIWZIQ



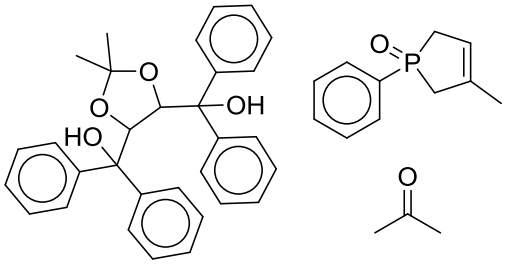
SUDJOS



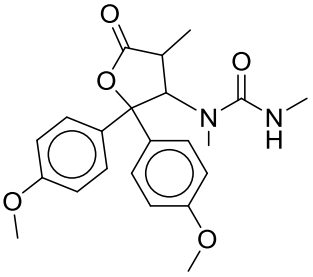
CANPUC



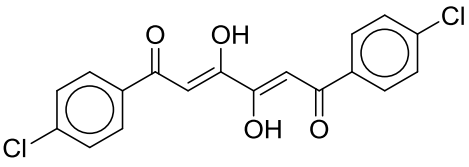
POFQAE



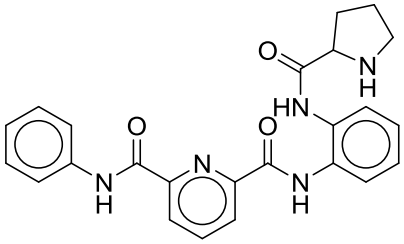
HOXBSA01



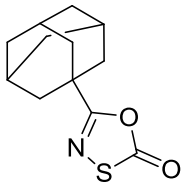
HIBBUQ



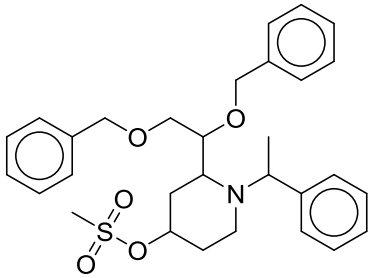
IQANUL



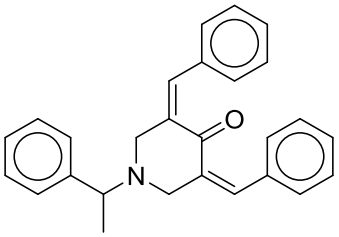
JUJTIR



RULJAK

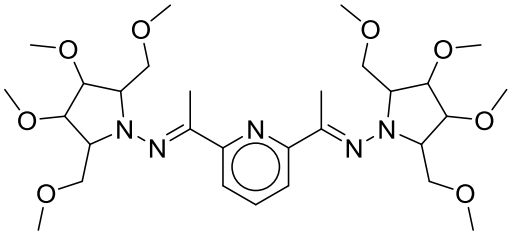


HEMYON

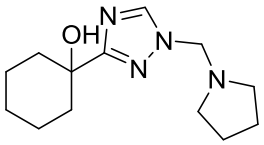


KUKBUO

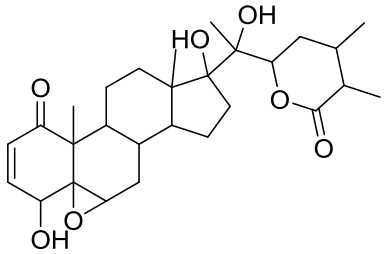
CIDFAX



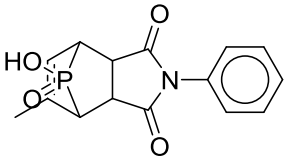
HAWDIT



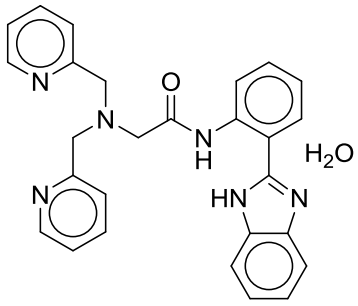
SEMDET



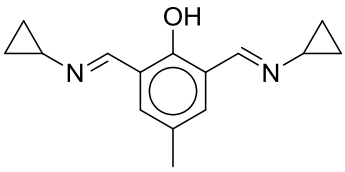
XIVYEG



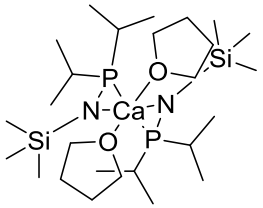
LIGTIE



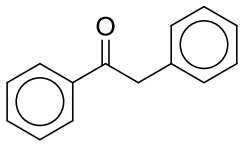
FIHYIG



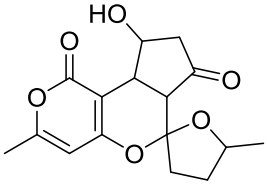
LIWGAA



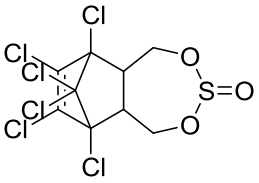
QAWQEM



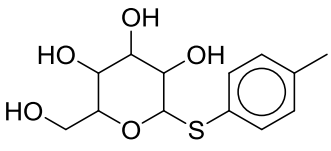
JULJEF



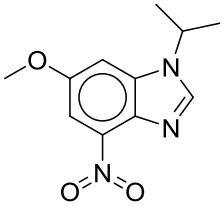
VAWMUD



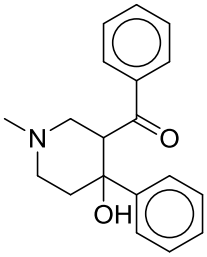
ENSULF03

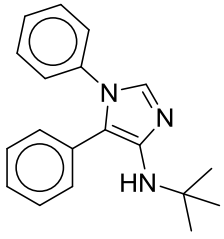


KAKRIA

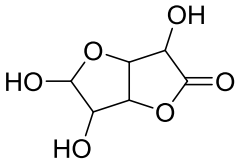


JOFGER

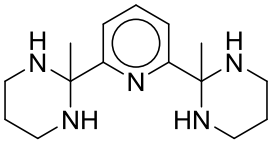




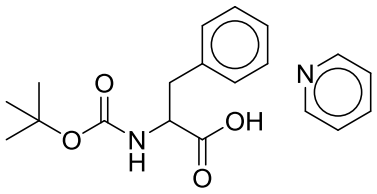
WOKJOV



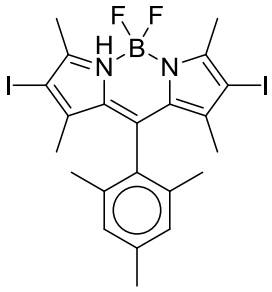
ZUGQUN



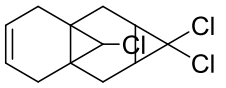
YOSRAZ



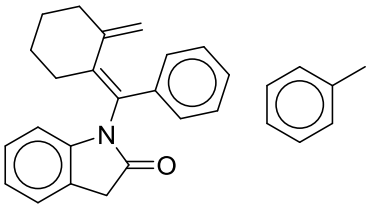
QAJPOI



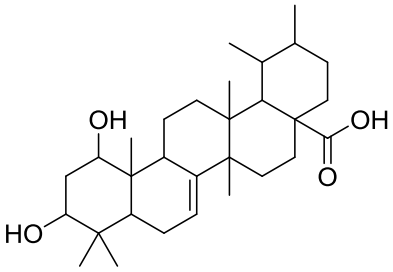
RUJCII01



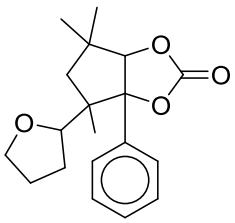
OQERAE01



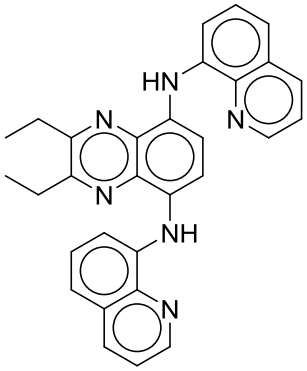
TAMTEG



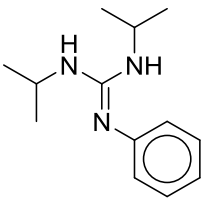
QOBSOQ



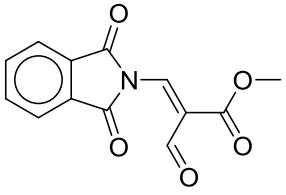
FARQIZ



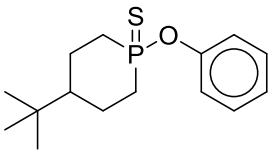
JANYON



TUKSIC

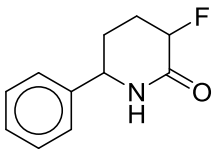


IPAPUM

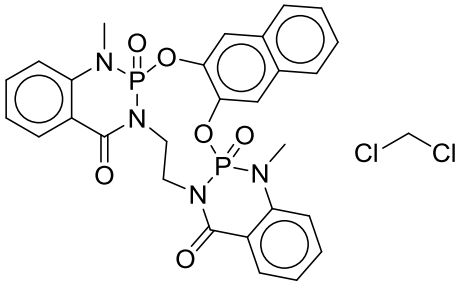


DOMSUT

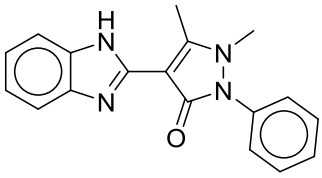
ZUPFUL



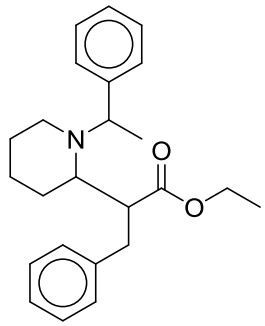
DOMJUM



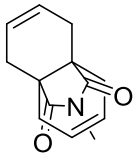
VABBUW



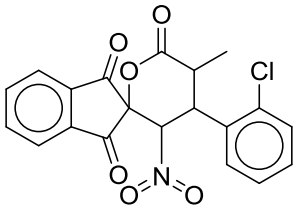
BABBOX



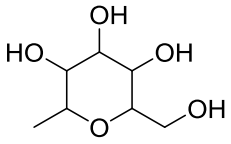
GIQLIC



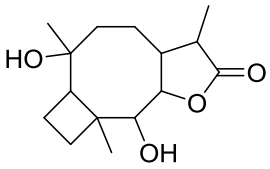
MZPROD



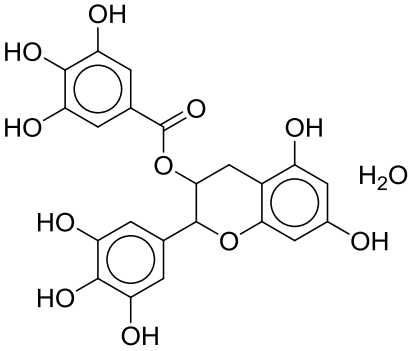
ZILGAE



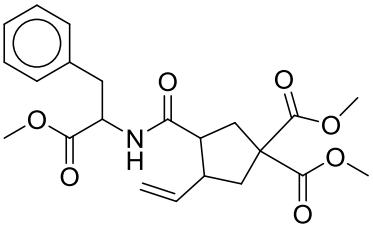
CAJNAC



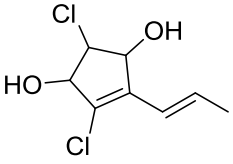
WOHRES



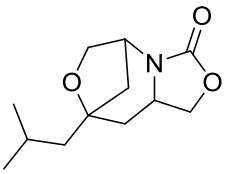
BONKOF



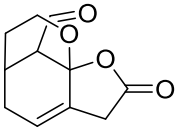
CEMCEE



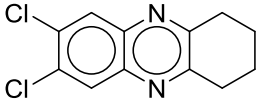
WADSIG



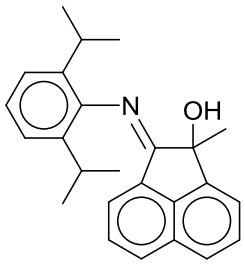
ULATAE



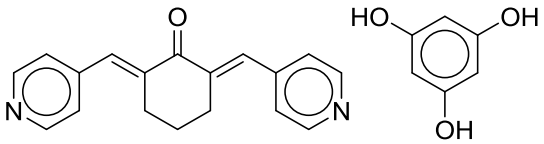
CISQIF



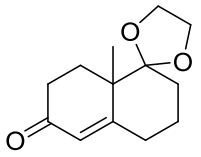
KIKSAZ



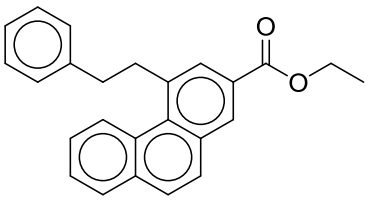
ZURCEW



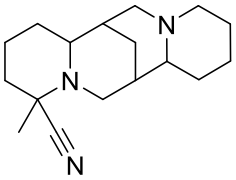
POLMAG



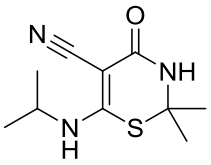
AZALAP



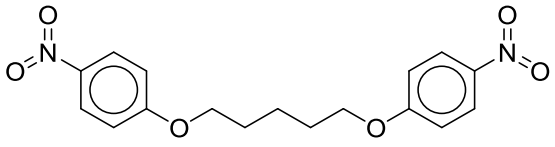
MERPUV



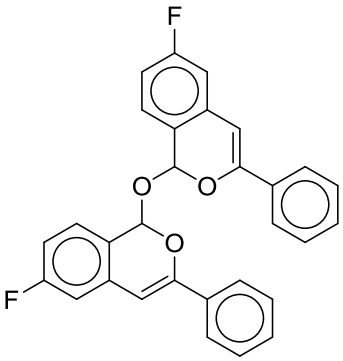
YUHYAB



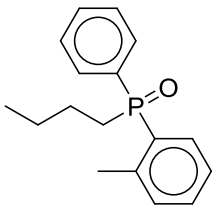
GIYZUK



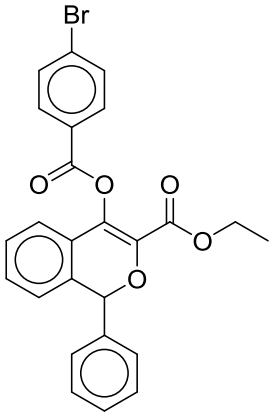
MOTFAD



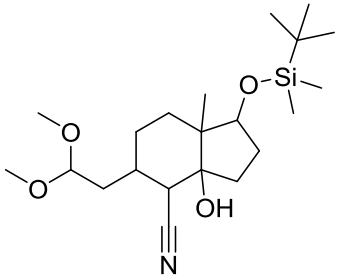
FABYAL

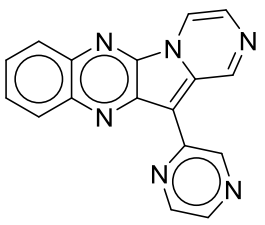


TOCVAK

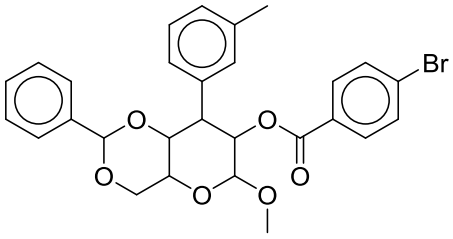


LAXROT



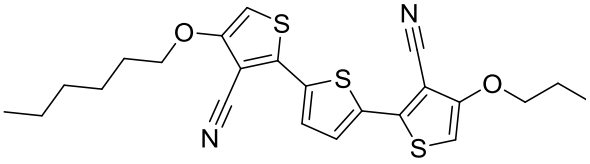


QEZBOM

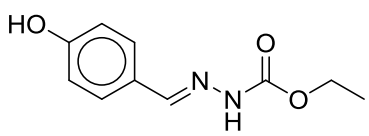


TABDIJ

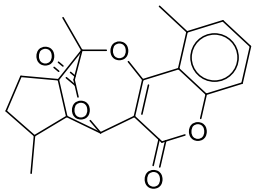
NAMVII



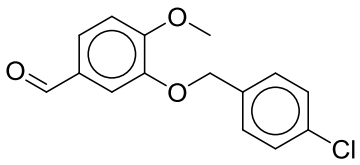
RELRAD



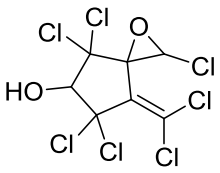
XOGXEX



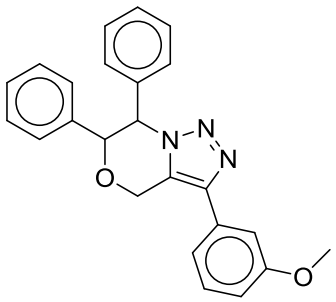
CUHDAM



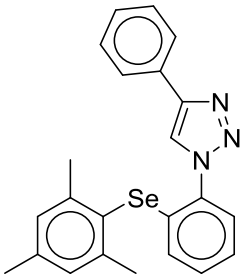
SETGEE



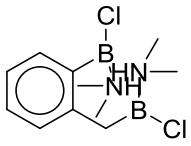
SEZMIT



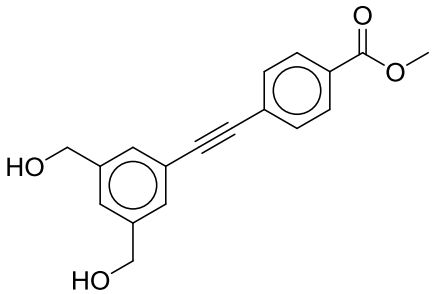
OGUCAV



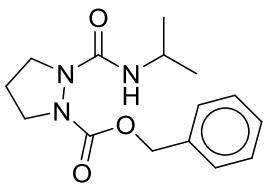
HUDJUN



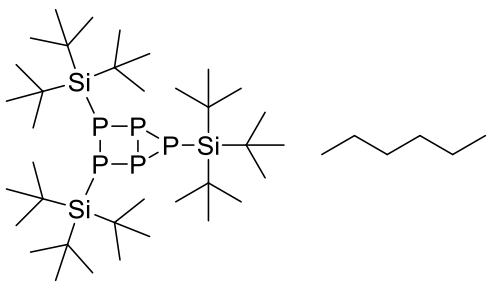
TOFHIF



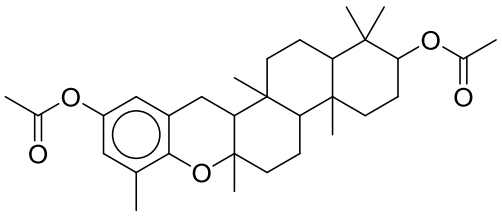
NENFIX01



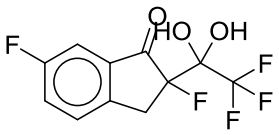
PILCOC



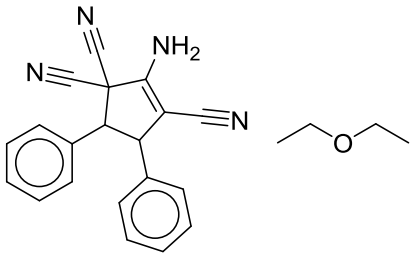
CAQJEK



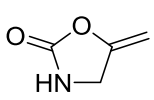
MAZHAY



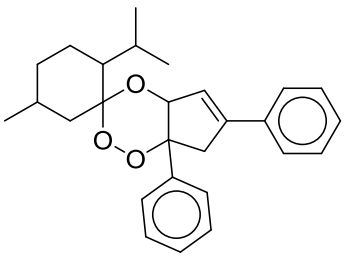
QUDWAP



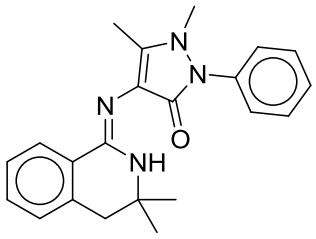
GOJXOS



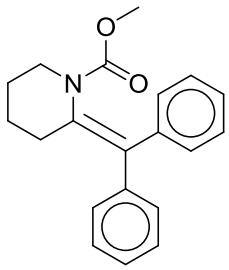
YOBQAH



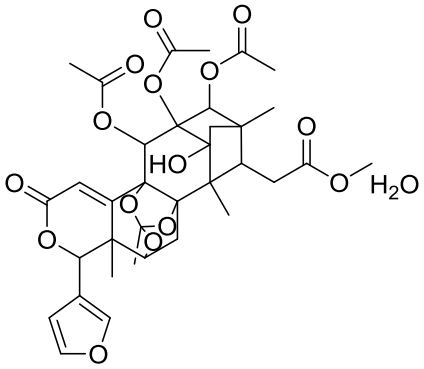
JOJCOA



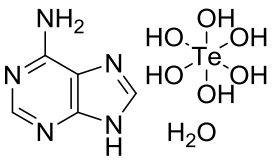
TUHWUO



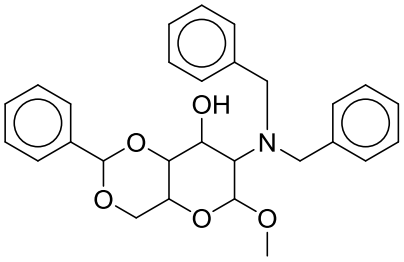
SAWVIX



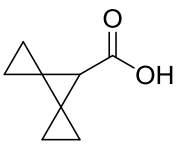
UWASUI



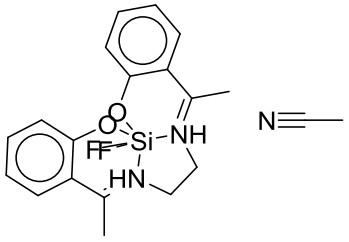
GUNQUB



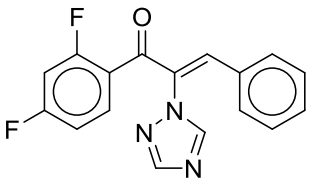
IMONON



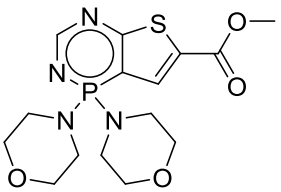
TEHMUO



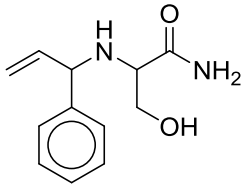
NUHLEH01



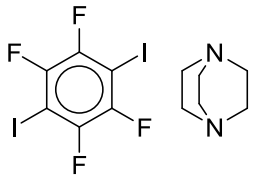
MUBKOK



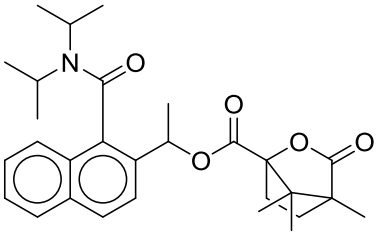
IDIBUT



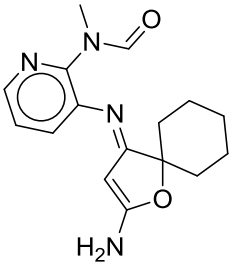
OJOPIN



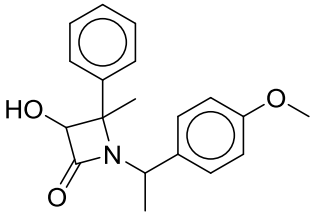
ISIHUN02



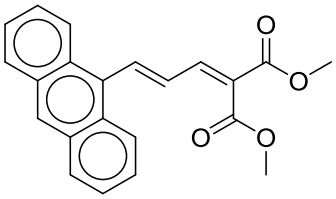
OFEYUT



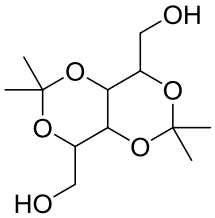
FEMXAY



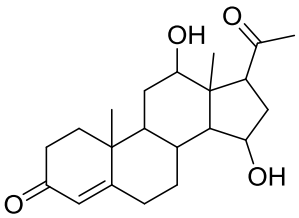
VAFVUU



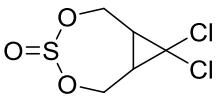
QIKYIU



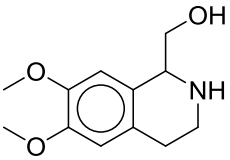
KOJYOX



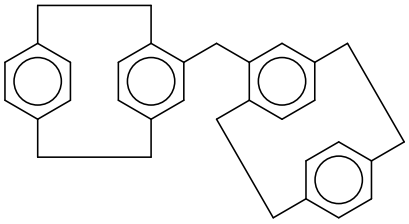
BAKVEQ



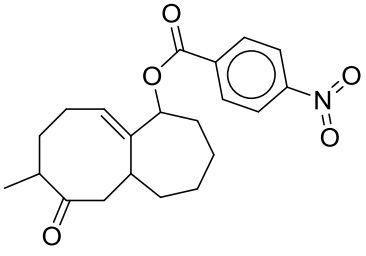
MOLHIF



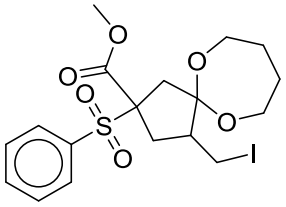
BIMCEG



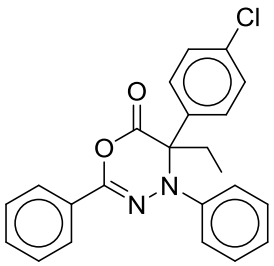
CUZYEB



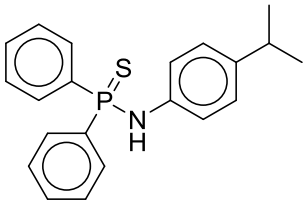
XIXBEM



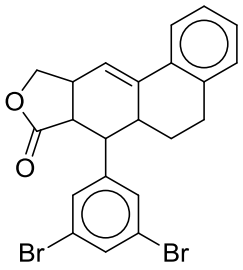
IYIWIW



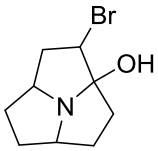
MORYAU



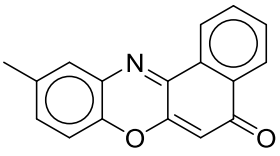
SULMIW



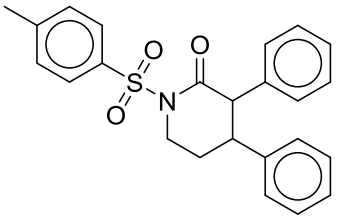
CAZMOH



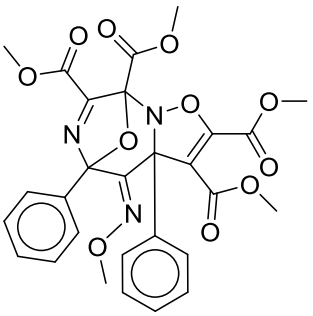
ACIQEI



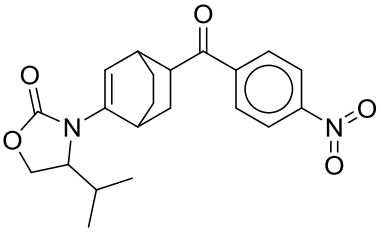
TUWJEC



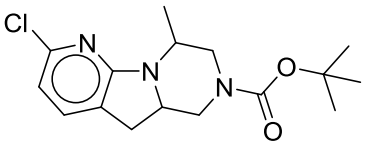
CEZMEB



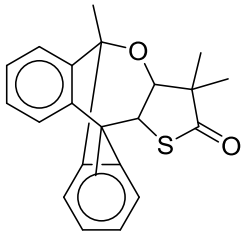
JADJEE



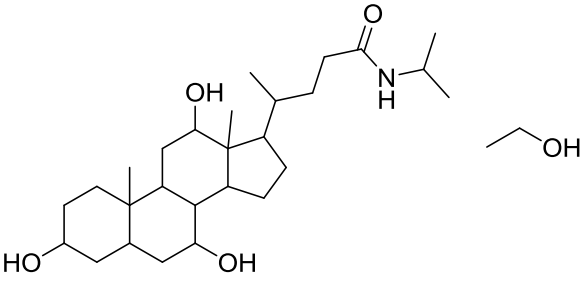
TIMYUL



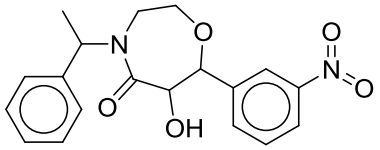
IFAXAP



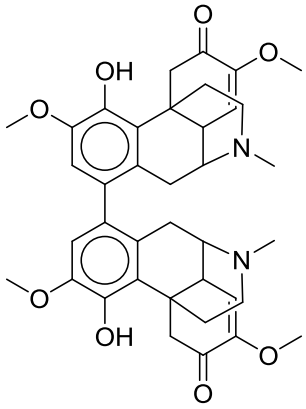
JUWNAQ



HEHQIU

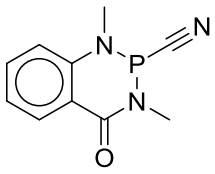


UBAWEB

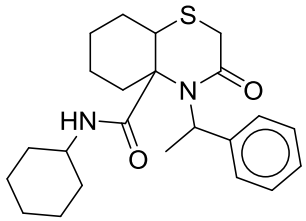


H_2O

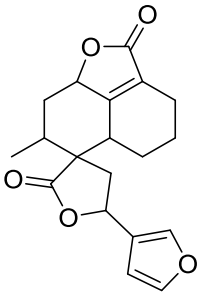
KAVZIS



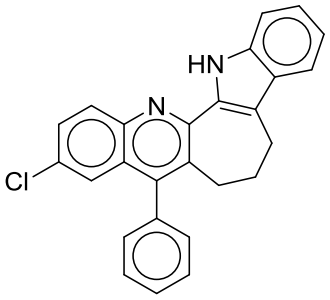
RIJJOJ



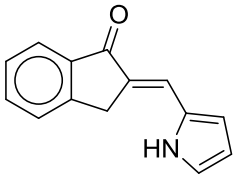
XUQMIF



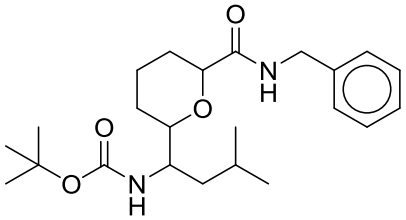
TEUVID



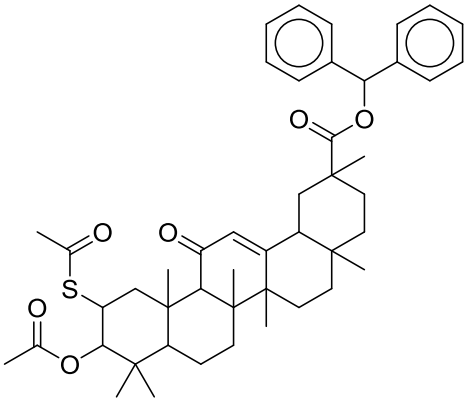
TUWYOA



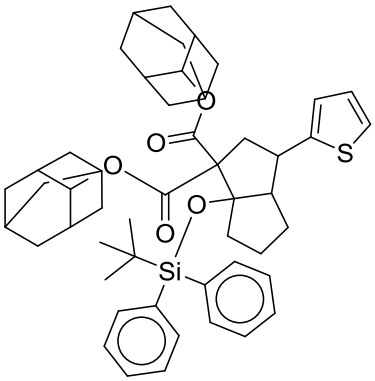
YABBIP



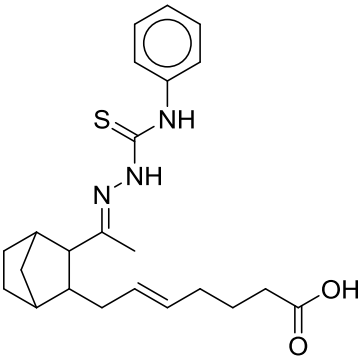
MELNAT



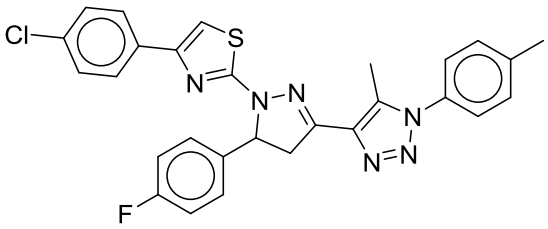
WUSPOQ



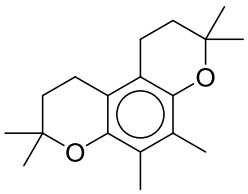
BEQZOO



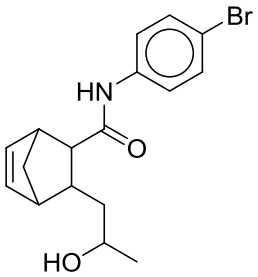
SEMVUB



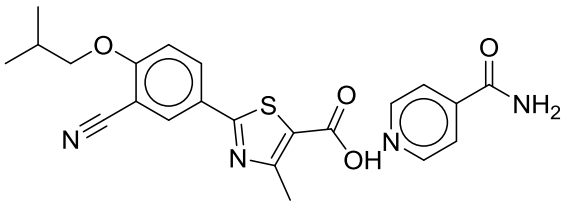
VEFFAP



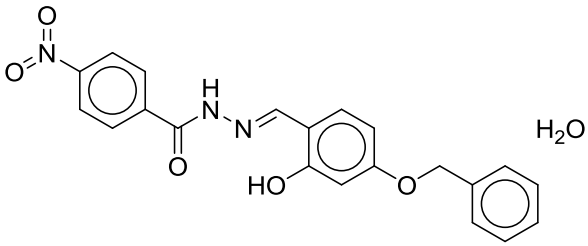
GOKSUV



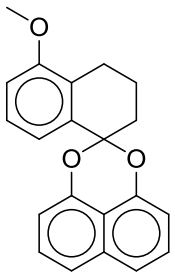
XIQJUD



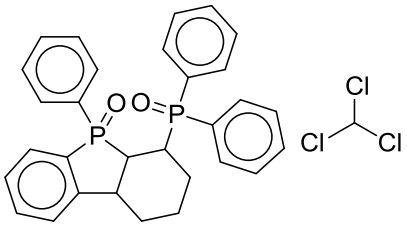
OYADAV



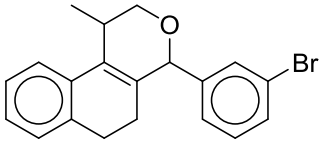
ZAPYIA



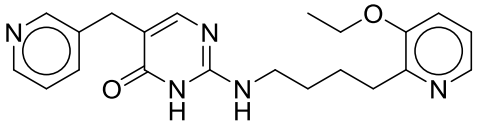
XIZTEF



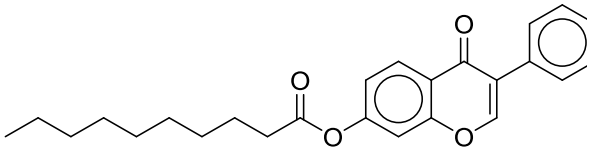
YEQCEC



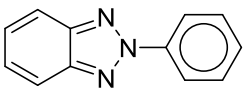
QOHMUX



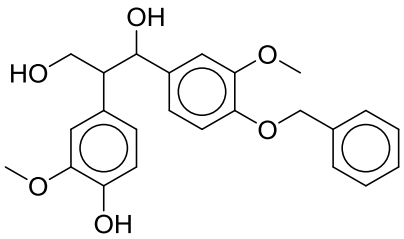
LEJMOC



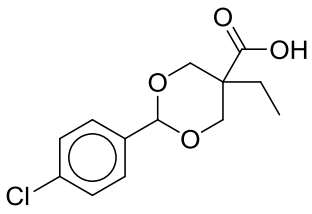
SABVAU



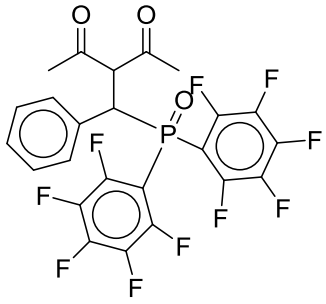
NIMMUS



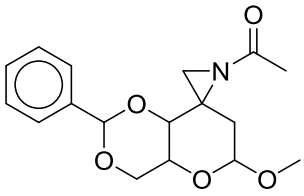
SIPJIK



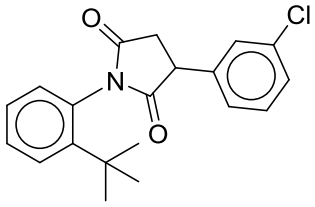
WUPQUV



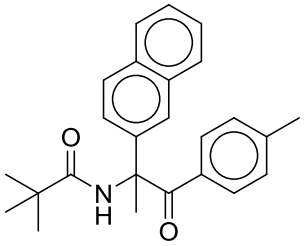
MOXYUV



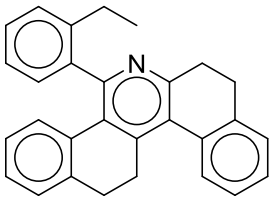
AZBAPY



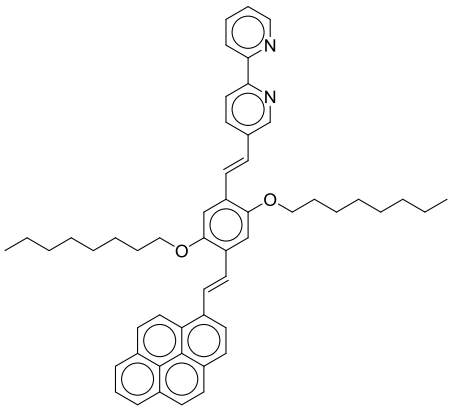
DIFYEX



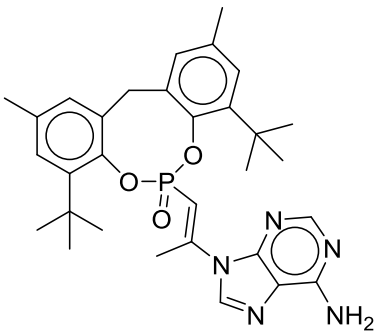
IMIRON



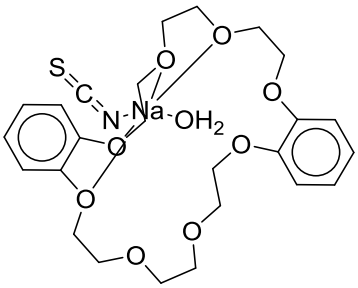
PUPFOW



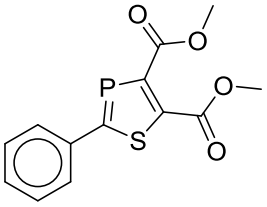
TIDNOK



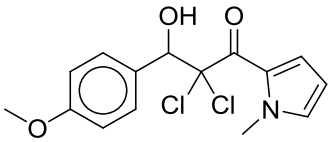
CEQYON



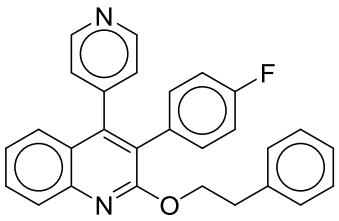
GEGZIC



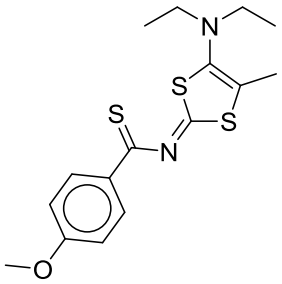
KUJDOI



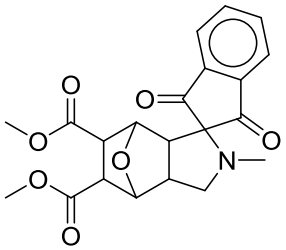
BETPEX



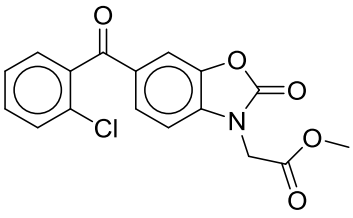
PIWWOI



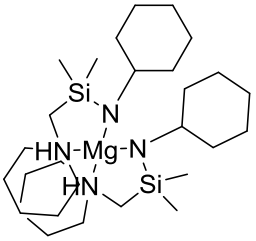
DASWOJ



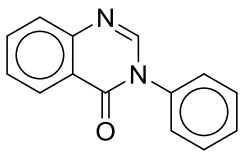
TAVJOR



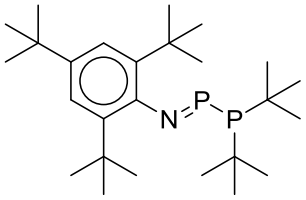
JEWCEU



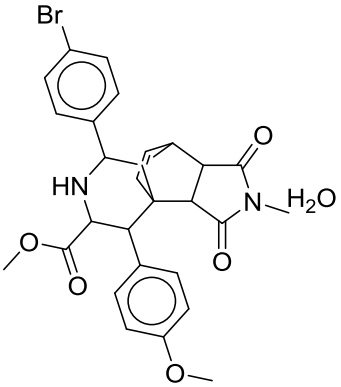
XIKZOI



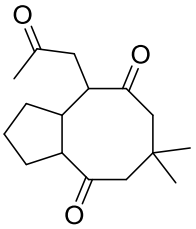
BAPBUQ



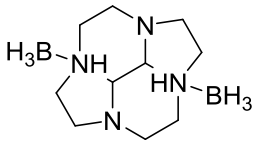
KIFTOI



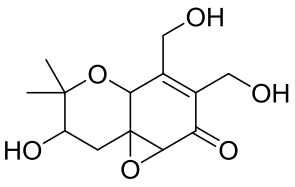
ZESVOK



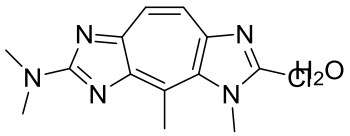
RASWEN



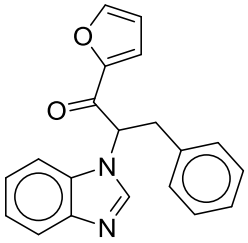
YAHTOQ



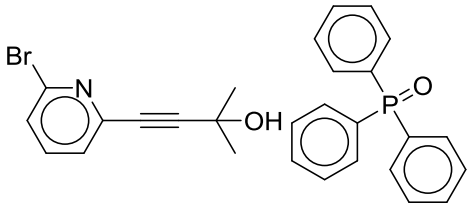
GILFOY



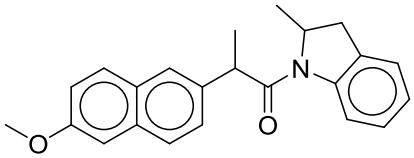
ZCPTAZ



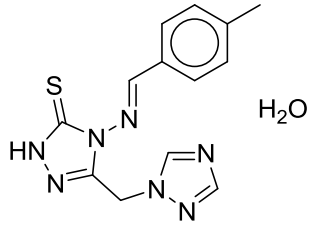
TIHTAG



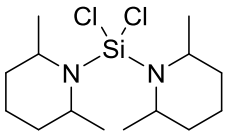
LUMYOH



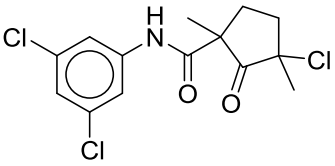
MOYKOA



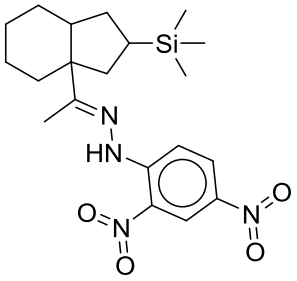
XAZMAN



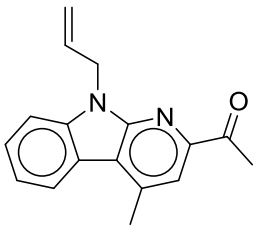
JOZGAG



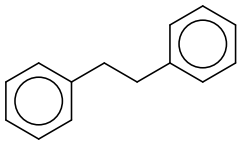
YOJWOL



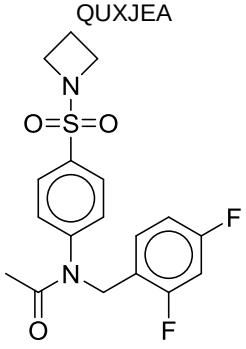
VEVZUR10



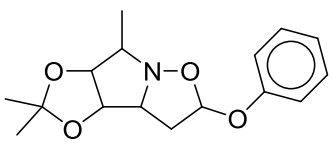
QUXJEA

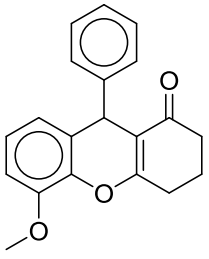


DIBENZ14

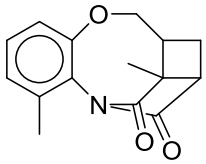


CEGFOL

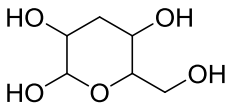




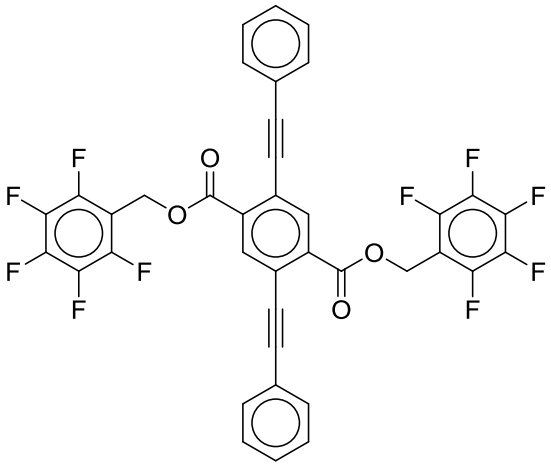
YABCUA



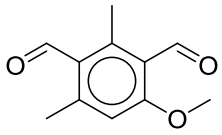
VAHBAK



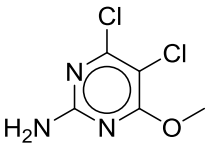
FUSJOU



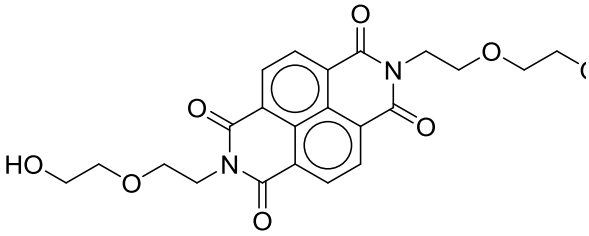
YIGGIF



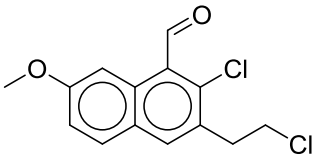
LOMTIS



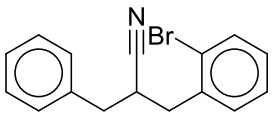
FATVUS



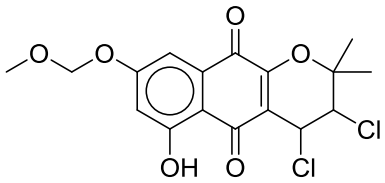
TUDYIB



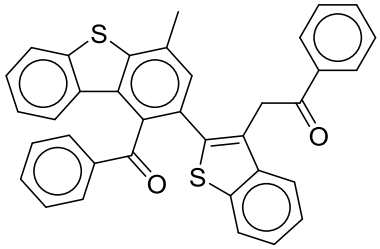
XOQLUK



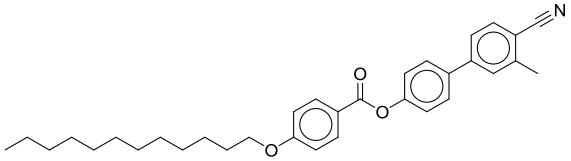
WOLWUQ



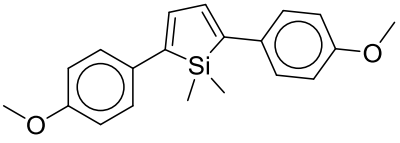
PEMFET



IGAXOE

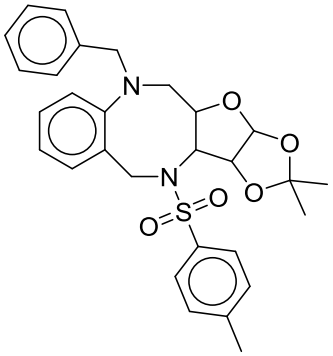


OPILUV

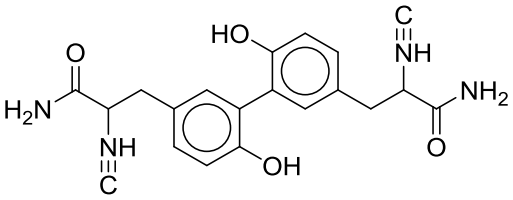


PUWROQ

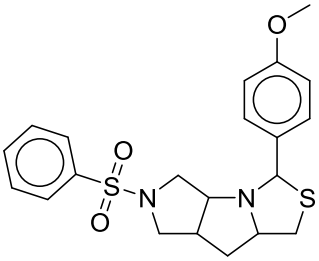
BAMKEF



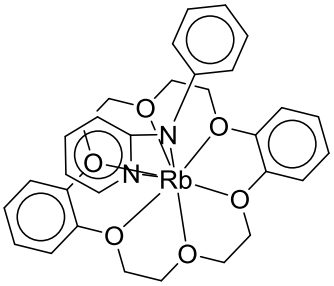
KUWXEG



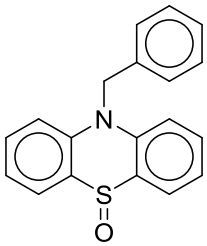
BIVFER



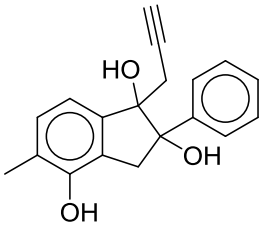
GITKEA



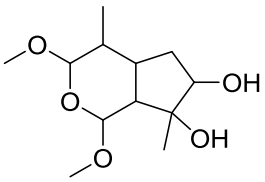
MAHRUJ



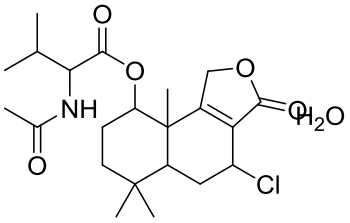
MUBSIM



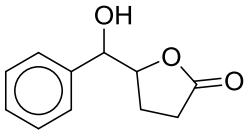
XUYRUF



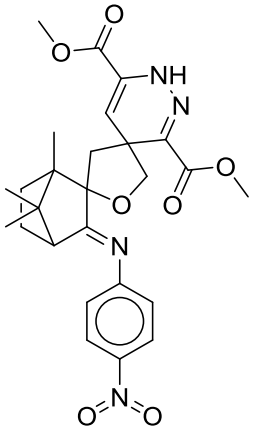
QULPEU



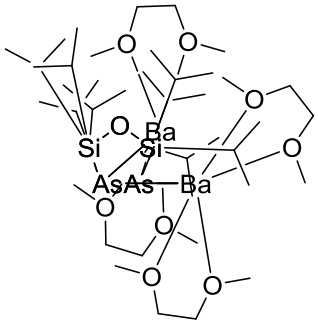
OYONUN



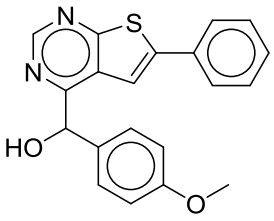
BALQUB



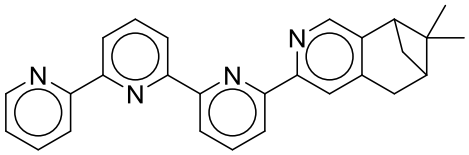
VIWGOY



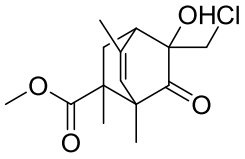
OHADEI



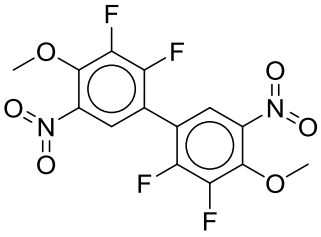
AZITEI



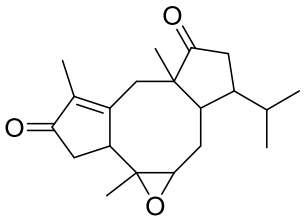
EKOJUL



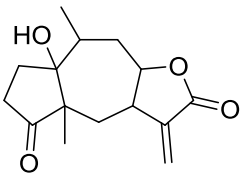
KEGWOJ



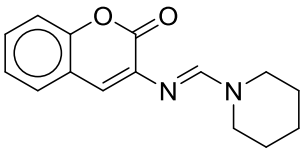
INOKEC



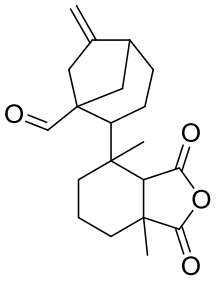
SIQHEF



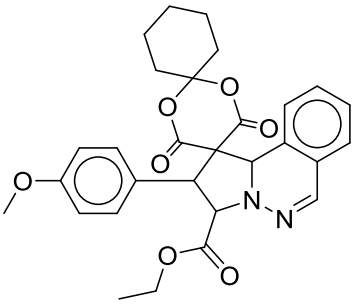
LUDGIA



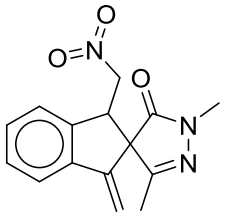
IQENIC



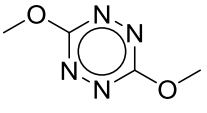
DEGBOG



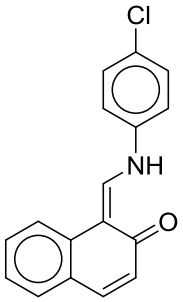
WEGYAJ



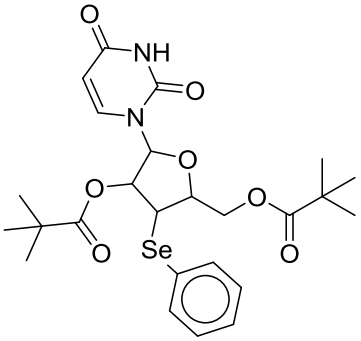
WAJREH



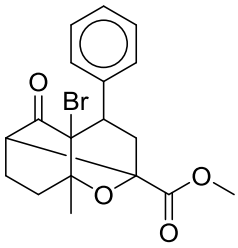
FIZJED



IHARAK

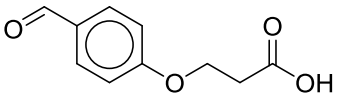


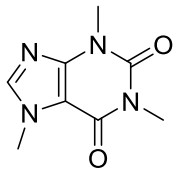
TANZEN



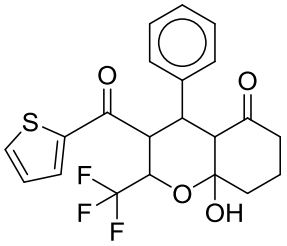
FUBWEG

YOPXUX

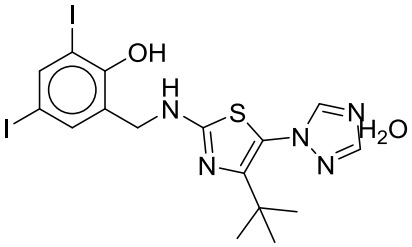




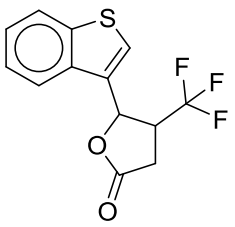
TENNOR



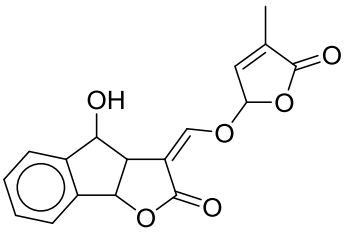
XICGEX



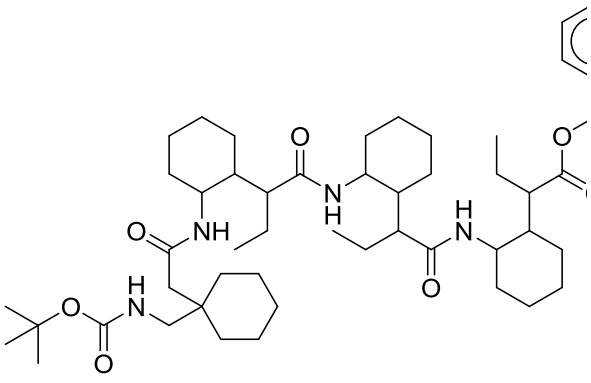
DORFIB



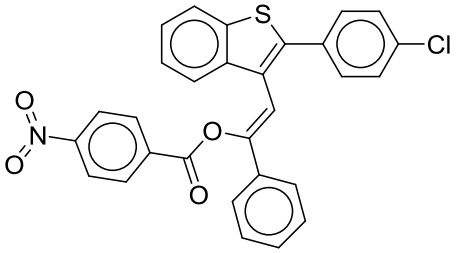
APEMOZ



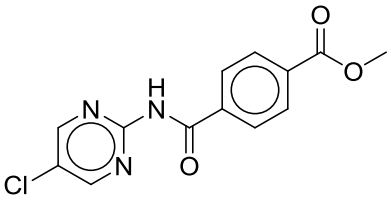
ANUTUZ



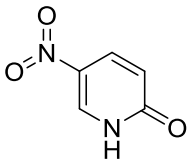
ISONIO



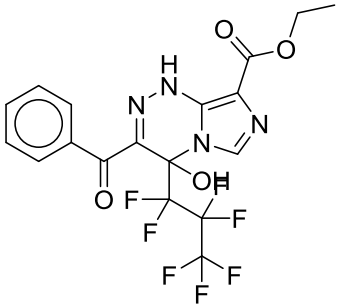
NEHMOE



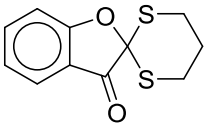
AWUGAB



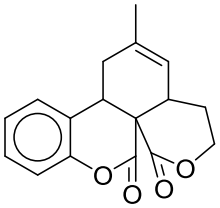
ELOXOS



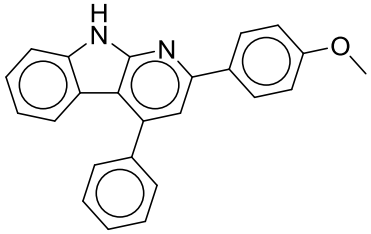
NAJQIA



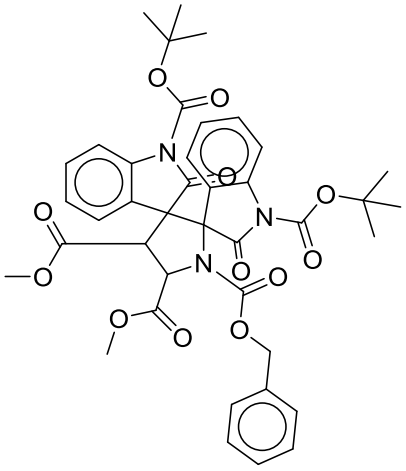
WAFKUM



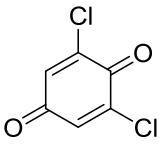
ISUSAS



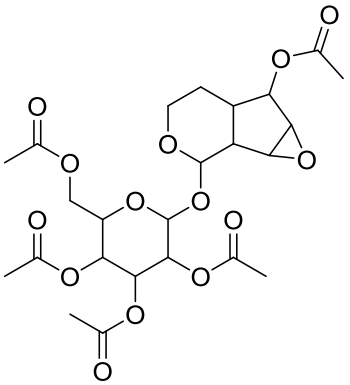
CANTOD



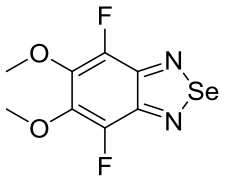
QUGKAG



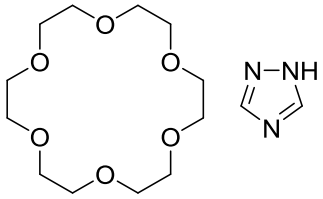
DCLBQN02



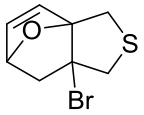
PULCHB11



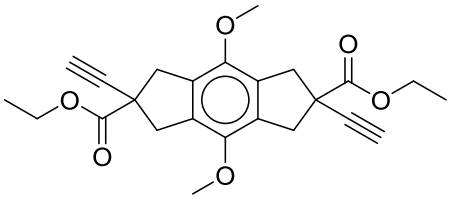
NABSER



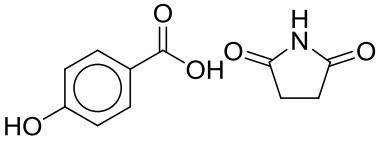
ZIJQOY



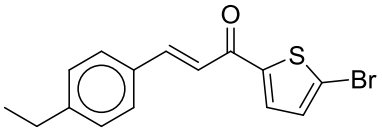
RAPJUO



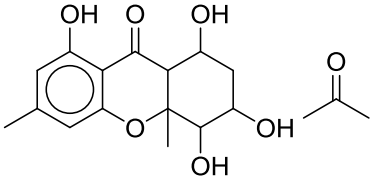
NIDWEE



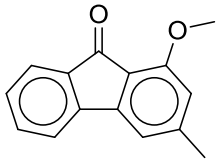
UGOTUH



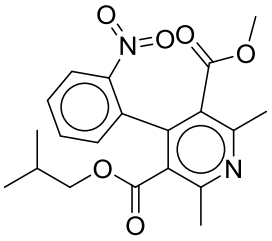
PUSKUL



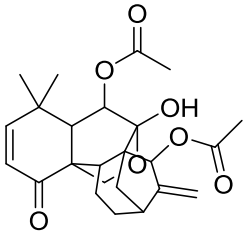
CEZMUQ



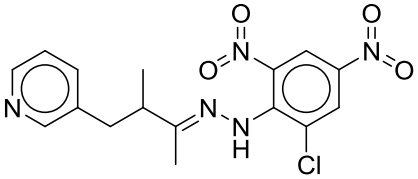
POZGOB



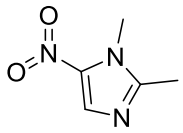
HUFKOJ



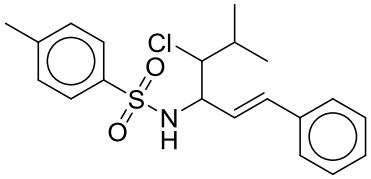
RAFYED



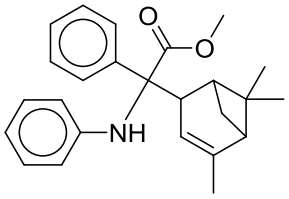
EDUJAO



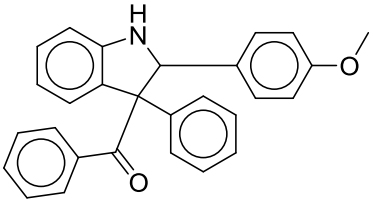
JIZJIL



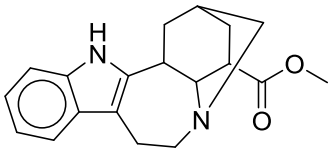
KEPPAY



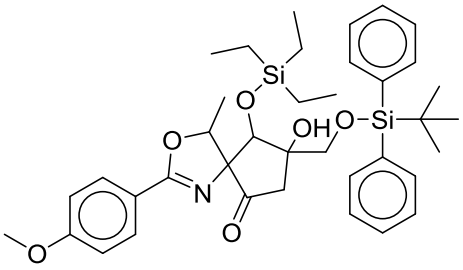
YABVUT



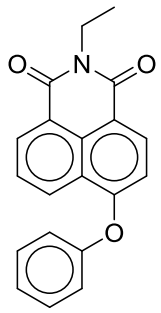
WUGSEY



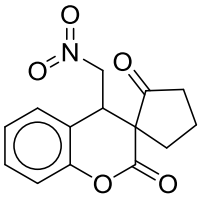
SAWRAL



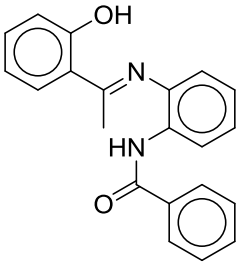
ONOYOG



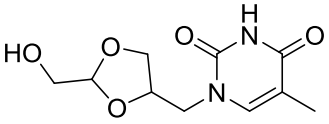
RICKEV



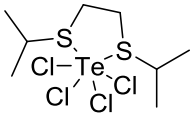
WEGPUV



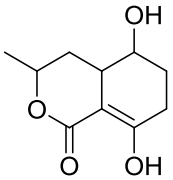
POSNAN



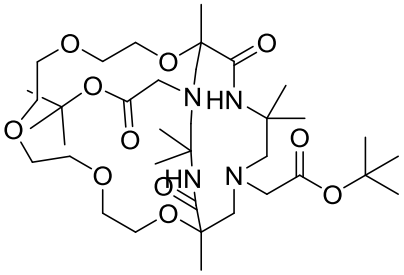
YOGYIC



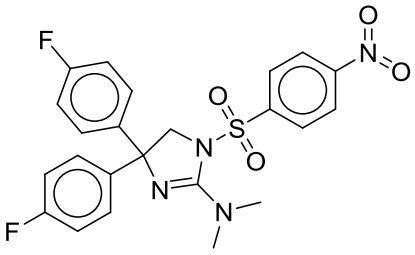
XUFDAE



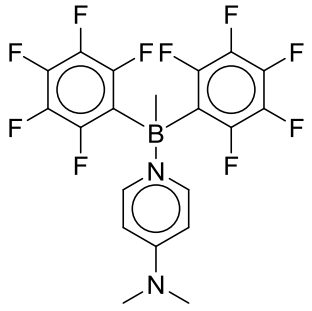
GUYBOR01



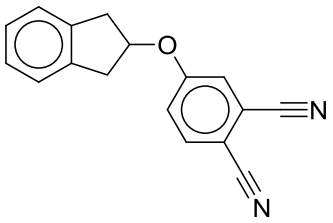
BEHXES



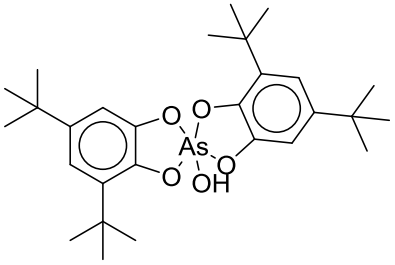
UDAHUE



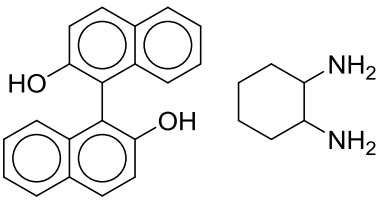
TUQSUV



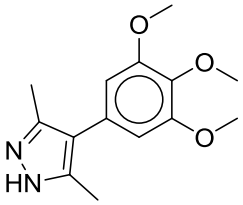
ILIREA



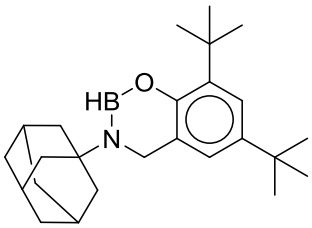
DADXIP



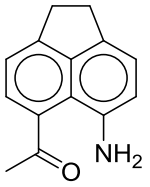
DIFFOO



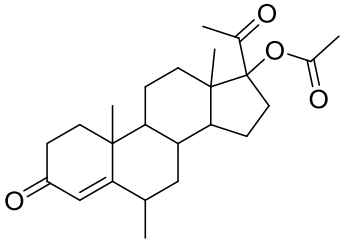
OKIKAW



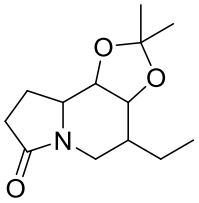
FUZXAB



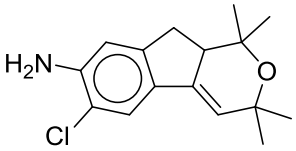
AYASEZ



YAWKAK



ACIVUF



OHULOU

References.

[1] IUPAC. Compendium of Chemical Terminology, 2nd ed. Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. https://doi.org/10.1351/goldbook.C01282, (accessed 21/07/18).
[2] IUPAC. Compendium of Chemical Terminology, 2nd ed. Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. https://doi.org/10.1351/goldbook.C01249, (accessed 21/07/18).
[3] IUPAC. Compendium of Chemical Terminology, 2nd ed. Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. https://doi.org/10.1351/goldbook.C01258, (accessed 21/07/18).
[4] M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11–15.
[5] M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870–2871.
[6] R. Wasylishen, T. Schaefer, *Can. J. Chem.*, 1972, **50**, 2710–2712.
[7] A. A. van Beuzekom, F. A. A. M. de Leeuw, C. Altona, *Magn. Reson. Chem.*, 1990, **28**, 68–74.
[8] P. E. Hansen, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1981, **14**, 175–295.
[9] M. Tafazzoli, M. Ghiasi, *Carbohydr. Res.*, 2007, **342**, 2086–2096.
[10] G. Palermo, R. Riccio and G. Bifulco, *J. Org. Chem.*, 2010, **75**, 1982–1991.
[11] T. Parella, F. Sánchez-Ferrando A. Virgili, *Magn. Reson. Chem.*, 1994, **32**, 657–664.
[12] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Cryst.,* 2016, B**72**, 171–179.
[13] L. M. Harwood, T. D. W. Claridge, *Introduction to organic spectroscopy*, Oxford University Press, 1997.
[14] W. Clegg, *Crystal structure determination*, Oxford University Press, 1998.
[15] C. D. Blundell, M. J. Packer, A. Almond, *Bioorg. Med. Chem.,* 2013, **21**, 4976–4987.
[16] C. Tormena, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2016, **96**, 73–88.
[17] G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, *Chem. Rev.*, 2007, **107**, 3744–3779.
[18] J. Wu, P. Lorenzo, S. Zhong, M. Ali, C. P. Butts, E. L. Myers, V. K. Aggarwal, *Nature*, 2017, **547**, 436–440.
[19] M. Burns, S. Essafi, J. R. Bame, S. P. Bull, M. P. Webster, S. Balieu, J. W. Dale, C. P. Butts, J. N. Harvey, V. K. Aggarwal, *Nature*, 2014, **513**, 183–188.
[20] J. Jeener, Lecture notes from Ampere Summer School, Basko Polje, Yugoslavia, 1971.
[21] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.*, 1976, **64**, 2229–2246.
[22] T. D. W. Claridge, *High-resolution NMR techniques in organic chemistry,* Vol. 27, Elsevier, 2008.
[23] K. C. Nicolaou, S. A. Snyder, *Angew. Chem. Int. Ed.,* 2005, 44, 1012-1044.
[24] M. E. Maier, *Nat. Prod. Rep.*, 2009, 26, 1105-1124.
[25] N. Grimblat, A. M. Sarotti, *Chem. Eur. J.*, 2016, **22**, 12246–12261.
[26] F. Della-Felice, R. A. Pilli, A. M. Sarotti, *J. Braz. Chem. Soc.*, 2018, 29, 1041–1075.
[27] K. Ermanis, K. E. B. Parkes, T. Agback, J. M. Goodman, *Org. Biomol. Chem.*, 2016,**14**, 3943–3949.
[28] M. Elyashberg, A. Williams, K. Blinov, *Contemporary Computer-Assisted Approaches to Molecular Structure Elucidation*, The Royal Society of Chemistry, 2012.
[29] M. Elyashberg, A. J. Williams K. Blinov, *Nat. Prod. Rep.*, 2010, **27**, 1296–1328.
[30] Y. D. Smurnyy, M. E. Elyashberg, K. A. Blinov, B. A. Lefebvre, G. E. Martin, A. J. Williams, *Tetrahedron*, 2005, **61**, 9980-9989.
[31] E. Troche-Pesqueira, C. Anklin, R. R. Gil, A. Navarro-Vázquez, *Angew. Chem.* 2017, **129**, 3714–3718.
[32] A. Navarro-Vázquez, R. R. Gil, K. Blinov, *J. Nat. Prod.*, 2018, **81**, 203–210.
[33] M. H. Levitt, *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons, 2001.
[34] C. P. Slichter, *Principles of magnetic resonance,* Vol. 1, Springer-Verlag, 1978.
[35] H. Günther, *NMR spectroscopy: basic principles, concepts and applications in chemistry,* John Wiley & Sons, 1995.
[36] P. J. Hore, *Nuclear magnetic resonance*, Oxford University Press, 1995.
[37] R. K. Harris, *Nuclear magnetic resonance spectroscopy*, Longman Scientific and Technical, 1987.
[38] E. F. Mooney, *An introduction to 19F NMR spectroscopy*, Heyden, 1970.
[39] P. E. Hansen, J. Spanget-Larsen, *Molecules*, 2017, **22**, 552–573.
[40] M. Silva, *Molecules*, 2017, **22**, 247–268.

[41] T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.*, 2007, **2**, 2451–2458.
[42] J. Keeler, *Understanding NMR spectroscopy*, John Wiley & Sons, 2010.
[43] A. W. Douglas, M. Shapiro, *Org. Magn. Reson.*, 1980, **14**, 38–39.
[44] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99
[45] J. M. Goodman, W. C. Still, *J. Comput. Chem.*, 1991, **12**, 1110–1117.
[46] K. D. Gibson, H. A. Scheraga, *J. Comput. Chem.*,1987, 8, 826–834.
[47] A.R. Leach, K. Prout, D. P. Dolata, *J. Computer-Aided Mol. Des.,* 1988, **2**, 107–123.
[48] A.R. Leach, D. P. Dolata, K. Prout, *J. Chem. Inf. Model.*, 1990, **30**, 316–324.
[49] M. Saunders, *J. Am. Chem. Soc.*, 1987, **109**, 3150–3152.
[50] D. M. Ferguson, D. J. Raber, *J. Am. Chem. Soc.*, 1989, **111**, 4371–4378.
[51] Z. Q. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A*, 1987, **84**, 6611–6615.
[52] G. Chang, W. C. Guida, W. C. Still, *J. Am. Chem. Soc.*, 1989, **111**, 4379–4386.
[53] G. M. Crippen, *J. Comput. Chem.*, 1982, **3**, 471–476.
[54] P. Hohenberg, W. Kohn, *Phys. Rev. B,* 1964, **136**, 864–871.
[55] A. R. Leach, *Molecular modelling: principles and applications,* Pearson Education, 2001.
[56] W. Kohn, A. D. Becke, R. G. Parr, *J. Phys. Chem.*, 1996, **100,** 12974–12980.
[57] W. Kohn, L. J. Sham, *Phys. Rev. A,* 1965, **140**, 1133–1138.
[58] D. A. McQuarrie, J. D. Simon, *Physical Chemistry, a molecular approach*, University Science Books, 1997.
[59] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.*,1984, **80**, 3265-3269.
[60] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294-301.
[61] J. S. Binkley, J. A. Pople, W. J. Hehre, *J. Am. Chem. Soc.*, 1980, **102**, 939–947.
[62] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.
[63] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
[64] P. C. Hariharan, J. A. Pople, *Theoret. Chim. Acta*, 1973, **28**, 213–222.
[65] P. C. Hariharan, J. A. Pople, *Mol. Phys.*, 1974, 27, 209–214.
[66] M. S. Gordon, Chem. *Phys. Lett.*, 1980, **76**, 163–168.
[67] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople *J. Chem. Phys.*, 1982, **77**, 3654–3665.
[68] R. C. Binning Jr., L. A. Curtiss, *Comput. Chem.*, 1990, **11**, 1206–1216.
[69] A. D. McLean, G. S. Chandler, *J. Chem. Phys.*, 1980, **72**, 5639–5648.
[70] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys*., 1980, **72**, 650–654.
[71] J. A. Pople, W. J. Hehre, *J. Comp. Phys.*, 1978, **27**, 161–168.
[72] F. Jensen, *J. Chem. Phys*., 2001, **115**, 9113–9125.
[73] F. Jensen, *J. Chem. Theory Comput*., 2006, **2**, 1360–1369.
[74] F. Jensen, *Theor. Chem. Acc.*, 2010, **126**, 371–382.
[75] H. Kjær, S. P. A. Sauer, *J. Chem. Theory Comput.*, 2011, **7**, 4070–4076.
[76] P. F. Provasi, G. A. Aucar, S. P. A. Sauer, *J. Chem. Phys.*, 2001, **115**, 1324–1334.
[77] P. F. Provasi, S. P. A. Sauer, *J. Chem. Phys.*, 2010 **133**, 1-10.
[78] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.
[79] W. Deng, J. R. Cheeseman, M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 1028–1037.
[80] A. D. Becke, *J. Chem. Phys*., 1993, **98**, 5648–5652.
[81] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
[82] C. Adamo, V. Barone, *J. Chem. Phys.*, 1998, **108**, 664–675.
[83] J. Chai, M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.
[84] J. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008,**10**, 6615–6620.
[85] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.

[86] N. F. Ramsey, *Phys. Rev.*, 1950, **77**, 567.
[87] N. F. Ramsey, *Phys. Rev.*, 1950, **78**, 699–703.
[88] N. F. Ramsey, *Phys. Rev.*, 1951, **83**, 540–541.
[89] N. F. Ramsey, *Phys. Rev.*, 1952, **86**, 243–246.
[90] N. F. Ramsey, *Phys. Rev.*, 1953, **91**, 303–307.
[91] N. F. Ramsey, *Phys. Rev.*, 1953, **89**, 527.
[92] G. Schreckenbach, T. Ziegler, *J. Phys. Chem.*, 1995, **99**, 606–611.
[93] J. R. Cheeseman, G. W. Trucks, T. A. Keith, M. J. Frisch, *J. Chem. Phys.*, 1996, **104**, 5497–5509.
[94] J. C. Facelli, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **58**, 176–201.
[95] N. F. Ramsey, E. M. Purcell, *Phys. Rev.*, 1952, **85**, 143–144.
[96] T. Helgaker, M. Watson, N. C. Handy, *J. Chem. Phys.*, 2000, **113**, 9402–9409.
[97] V. Sychrovský, J. Gräfenstein, D. Cremer, *J. Chem. Phys.*, 2000, **113**, 3530–3547.
[98] V. Barone, J. E. Peralta, R. H. Contreras, J. P. Snyder, *J. Phys. Chem. A*, 2002, **106**, 5607–5612.
[99] J. E. Peralta, G. E. Scuseria, J. R. Cheeseman, M. J. Frisch, *Chem. Phys. Lett.*, 2003, **375**, 451–458.
[100] S. Koide, E. Duval, *J. Chem. Phys.*, 1964, **41**, 315–320.
[101] W. Pauli, *Z. Phys.*, 1925, **31**, 765–783.
[102] F. Hund, *Z. Phys.*, 1925, **33**, 855–859.
[103] D. B. Davies, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1978, **12**, 135–225.
[104] J. Schmidt, *J. Biomol. NMR*, 2007, **37**, 287–301.
[105] V.F. Bystrov, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1976, **10**, 41–82.
[106] P. Salvador, *Annu*. *Rep*. *NMR Spectrosc*., 2014, **81**, 185–227.
[107] B. Coxon, *Adv. Carbohydr. Chem. Biochem.*, 2009, **62**, 17–82.
[108] C.A.G. Haasnoot, F.A.A.M. de Leeuw, C. Altona, *Tetrahedron*, 1980, **36**, 2783-2792.
[109] M. L. Huggins, *J. Am. Chem. Soc.*, 1953, **75**, 4123–4126.
[110] K. Imai, E. Ōsawa, *Magn. Reson. Chem.*, 1990, **28**, 668–674.
[111] E. Ōsawa, T. Ouchi, N. Saito, M. Yamato, O. S. Lee M. Seo, *Magn. Reson. Chem.*, 1992, **30**, 1104–1110.
[112] R. H. Contreras, J. E. Peralta, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2000, **37**, 321–425.
[113] N. Muller, D. E. Pritchard, *J. Chem. Phys.*, 1959, **31**, 768–771.
[114] J. N. Shoolery, *J. Chem. Phys.*, 1959, **31**, 1427–1428.
[115] A. S. Perlin, B. Casu, *Tetrahedron Lett.*, 1969, **10**, 2921–2924.
[116] S. Wolfe, B. M. Pinto, V. Varma, R. Y. N. Leung, *Can. J. Chem.*, 1990, **68**, 1051–1062,
[117] K. Bock, L. Wiebe, *Acta Chem. Scand.*, 1973, **27**, 2676–2678.
[118] G. J. Karabatsos, J. D. Graham, F. Vane, *J. Phys. Chem.*, 1961, **65**, 1657–1658.
[119] R. M. Lynden-Bell, N. Sheppard, *Proc. R. Soc. Lond. A*, 1962, **269**, 385–403.
[120] T. Leman, Master's thesis, University of Bristol, 2017.
[121] G. J. Karabatsos, C. E. Orzech Jr., N. Hsi, *J. Am. Chem. Soc.*, 1966, **88**, 1817–1818.
[122] R. U. Lemieux, T. L. Nagabhushan, B. Pau, *Can. J. Chem.*, 1972, **20**, 773–776.
[123] R. Aydin, H. Günther, *Magn. Reson. Chem.*, 1990, **28**, 448–457.
[124] A. Navarro-Vázquez, R. Santamaría-Fernández, F. J, Sardina, *Magn. Reson. Chem.*, 2018, **56**, 505–512.
[125] C. L. Dickson, C. D. Blundell, C. P. Butts, A. Felton, A. Jeffreys, Z. Takacs, *Analyst*, 2017,**142**, 621–633.
[126] M. Murata, S. Matsuoka, N. Matsumori, G. K. Paul, K. Tachibana, *J. Am. Chem. Soc*., 1999, **121**, 870–871.
[127] G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, *Chem. Rev.*, 2007, **107**, 3744–3779.
[128] T. Parella and J. F. Espinosa, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 17–55.
[129] B. L. Marquez, W. H. Gerwick, R. T. Williamson, *Magn. Reson. Chem.*, 2001, **39**, 499–530.
[130] L. Castañar, T. Parella, *Annu. Rep. NMR Spectrosc.*, 2015, 84, 163–232.
[131] S. Gil, J. F. Espinosa, T. Parella, *J. Magn. Reson.*, 2011, **213**, 145–150.
[132] L. Castañar, J. Saurí, P. Nolis, A. Virgili, T. Parella, *J. Magn. Reson.*, 2014, **238**, 63–69.
[133] M. Misiak, W. Koźmiński, *Magn. Reson. Chem.*, 2009, **47**, 205–209.
[134] C. P. Butts, B. Heise, G. Tatolo, *Org. Lett.*, 2012, **14**, 3256–3259.
[135] G. K. Pierens, T. K. Venkatachalam, D. C. Reutens, *Magn. Reson. Chem.*, 2016, **54**, 941–946.
[136] A. G. Kutateladze, O. A. Mukhina, *J. Org. Chem.*, 2015, **80,** 10838–10848.
[137] R. A. E. Edden, J. Keeler, *J. Magn. Reson.*, 2004, 166, 53–68.
[138] S. Gil, J. F. Espinosa and T. Parella, *J. Magn. Reson.*, 2010, **207**, 312–321.
[139] M. Kurz, P. Schmieder and H. Kessler, *Angew. Chem., Int. Ed. Engl.*, 1991, **30**, 1329–1331.

[140] W. Koźmiński and D. Nanz, *J. Magn. Reson.*, 1997, **124**, 383–392.
[141] P. Nolis, T. Parella, *J. Magn. Reson.*, 2005, **176**, 15–26.
[142] W. Koźmiński, *J. Magn. Reson.*, 1999, **137**, 408–412.
[143] J. Saurí, J. F. Espinosa, T. Parella, *Angew. Chem., Int. Ed.*, 2012, **51**, 3919–3922.
[144] J. Saurí, T. Parella, *Magn. Reson. Chem.*, 2012, **50**, 717–721.
[145] MestReNova, Version 9.0.1, Mestrelab Research S.L., Santiago de Compostela, Spain, 2014.
[146] J. R. de Laeter, J. K. Böhlke, P. De Bièvre, H. Hidaka, H. S. Peiser, K. J. R. Rosman, P. D. P. Taylor, Pure App. Chem., 2003, **75**, 683–800.
[147] A. Jeffreys, Undergraduate thesis, University of Bristol, 2016.
[148] Z. Takacs, C4X discovery, private communication, 2016.
[149] S. A. Smith, T. O. Levante, B. H. Meier, R. R. Ernst, *J. Magn. Reson., Ser. A*, 1994, **106**, 75–105.
[150] M. Hatlo, F. James, P. Mato, L. Moneta, M. Winkler, A. Zsenei, *IEEE Trans. Nucl. Sci.*, 2005, **52**, 2818–2822.
[151] V. V. Krishnamurthy, *J. Magn. Reson., Ser. A*, 1996, **121**, 33–41.
[152] A. Meissner, O. W. Sørensen, *Magn. Reson. Chem.*, 2001, **39**, 49–52.
[153] K. Furihata, M. Tashiro and H. Seto, *Magn. Reson. Chem.*, 2010, **48**, 179–183.
[154] J. Saurí, T. Parella, *Magn. Reson. Chem.*, 2013, **51**, 509–516.
[155] R. T. Williamson, B. L. Marquez, W. H. Gerwick, K. E. Kövér, *Magn. Reson. Chem.*, 2000, **38**, 265–273.
[156] S. Boros, K. E. Kövér, *Magn. Reson. Chem.*, 2011, **49**, 106–110.
[157] L. Castañar, J. Saurí, R. T. Williamson, A. Virgili, T. Parella, *Angew. Chem., Int. Ed.*, 2014, **53**, 8379–8382.
[158] I. Timári, L. Szilágyi, K. E. Kövér, *Chem. Eur. J.*, 2015, **21**, 13939–13942.
[159] S. Wimperis, R. Freeman, *J. Magn. Reson.*, 1984, **58**, 348–353.
[160] C. Emetarom, T. L. Hwang, G. Mackin, A. J. Shaka, *J. Magn. Reson., Ser. A*, 1995, **115**, 137–140.
[161] L. Castañar, R. Roldán, P. Clapés, A. Virgili, T. Parella, *Chem. Eur. J.*, 2015, **21**, 7682–7685.
[162] K. Zangger, I. M. Armitage, *Magn. Reson. Chem.*, 2000, **38**, 452–458.
[163] C. E. Hadden, D. T. Angwin, *Magn. Reson. Chem.*, 2001, **39**, 1–8.
[164] J. San Fabián, J. M. García de la Vega, R. Suardíaz, M. Fernández-Oliva, C. Pérez, R. Crespo-Otero, R. H. Contreras, *Magn. Reson. Chem.*, 2013, **51**, 775–787.
[165] A. S. Edison, J. L. Markley, F. Weinhold, *J. Phys. Chem.*, 1993, **97**, 11657–11665.
[166] V. Blechta, F. del Río-Portilla, R. Freeman, *Magn. Reson. Chem.*, 1994, **32**, 134–137.
[167] J. San Fabián, J. Guilleme, E. Diez, *Journal of Molecular Structure: THEOCHEM*, 1998, 426, 117–133.
[168] Matlab R2016a (9.0.0.341360), MathWorks, 2016.
[169] R. E. Wasylishen, T. Schaefer, *Can. J. Chem.*, 1974, **52**, 3247–3250.
[170] R. Wasylishen, T. Schaefer, *Can. J. Chem.*, 1973, **51**, 961–973.
[171] R. V. Dubs, W. Von Philipsborn, *Org. Magn. Reson.*, 1979, **12**, 326–330.
[172] T. Parella, F. Sánchez-Ferrando, A. Virgili, *Magn. Reson. Chem.*, 1997, **35**, 30–34.
[173] J. B. Houseknecht, T. L. Lowary, C. M. Hadad, *J. Phys. Chem. A*, 2003, **107**, 372–378.
[174] C. P. Butts, C. R. Jones, J. N. Harvey, *Chem. Commun.*, 2011, **47**, 1193–1195.
[175] Schrödinger Release 2011-1: MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2011
[176] G. Chang, W. C. Guida, W. C. Still, *J. Am. Chem. Soc.*, 1989, **111,** 4379–4386.
[177] GaussView, Version 5, R. Dennington, T. A. Keith, and J. M. Millam, Semichem Inc., Shawnee Mission, KS, 2016.
[178] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
[179] The Cambridge Crystallographic Data Centre (CCDC), https://ccdc.cam.ac.uk/, (accessed 05/01/2018).
[180] I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Cryst.*, 2002, B58, 389–397.
[181] J. Clements, work performed as part of an undergraduate summer project at the University of Bristol, 2017.
[182] W. Gerrard, PhD student, University of Bristol, 2017.
[183] E. Johnston, Undergraduate thesis, University of Bristol, 2018.
[184] This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol, http://www.bris.ac.uk/acrc/.
[185] R. T. Williamson, A. V. Buevich, G. E. Martin, *Org. Lett.*, 2012, **14**, 5098–5101.

[186] G. E. Martin and J. Saurí, Merck, USA, private communication, 2017.
[187] S. N. Maximoff, J. E. Peralta, V. Barone, G. E. Scuseria, *J. Chem. Theory Comput.*, 2005, **1** 541–545.
[188] T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007–1023.
[189] R. A. Kendall, T. H. Dunning Jr., R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
[190] E. R. Davidson, *Chem. Phys. Lett.*, 1996, **260**, 514–518.
[191] D. M. Silver, S. Wilson, W. C. Nieuwpoort, *Int. J. Quantum Chem.*, 1978, **14**, 635–639.
[192] Q. Zhang, W. Zhang, Y. Li, J. Wang, L. Zhang, T. Hou, *J. Cheminformatics*, 2012, **4**, 26.
[193] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.
[194] M. H. Levitt, R. Freeman and T. Frenkiel, *J. Magn. Reson.*, 1982, **47**, 328–330.
[195] Ē. Kupče, J. Boyd and I. D. Campbell, *J. Magn. Reson., Ser. B,* 1995, **106**, 300–303.
[196] J. A. Nelder, R. Mead, *Comput. J.*, 1965, **7**, 308–313.
[197] X. Han, V. Pozdina, C. Haridas, P. Misra, *J. Inf. Comput. Sci.,* 2006, 3.
[198] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.
[199] T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 720–729.
[200] P. Pyykkö, M. Atsumi, *Chem. Eur. J.*, 2009, **15**, 12770–12779.
[201] M. Karthikeyan, R. Vyas, *J. Cheminformatics*, 2016, **8**, 73.
[202] J. M. Robertson, *Acta Cryst.*, 1948, **1**, 101–109.
[203] K. F. Riley, M. P. Hobson, S. J. Bence, *Mathematical methods for physics and engineering: a comprehensive guide,* Cambridge university press, 2006.
[204] Stack exchange, https://math.stackexchange.com/q/47084, (accessed 24/05/18).